

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-265499

(43)Date of publication of application : 15.10.1993

(51)Int.Cl.

G10L 9/18
G10L 9/00

(21)Application number : 04-092258

(71)Applicant : SONY CORP

(22)Date of filing : 18.03.1992

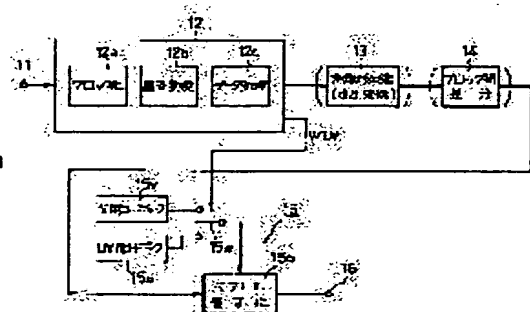
(72)Inventor : NISHIGUCHI MASAYUKI
MATSUMOTO ATSUSHI

(54) HIGH-EFFICIENCY ENCODING METHOD

(57)Abstract:

PURPOSE: To enable encoding of high quality by finding on-frequency-axis data in block units of an audio signal, and switching and quantizing plural code books corresponding to parameters showing features of the respective blocks.

CONSTITUTION: A frequency axis conversion processing part 12 provided with a block division part 12a which divides the input speech signal, etc., on the time base into blocks of a specific number of samples, an orthogonal conversion part 12b, a data processing part 12c for finding amplitude information representing the features of a spectrum envelope, etc., converts the speech signal, etc., into the spectrum amplitude data on the frequency axis. A vector quantization part 15 puts a specific number of input data into M-dimensional vectors and thus performs vector quantization. This vector quantization part 15 is equipped with plural kind of code books and switches the code books according to the parameters showing the features of the blocks of the input signal to quantize the signal. Consequently, the vector quantization is performed efficiently corresponding to the characteristics, etc., of the input data.



LEGAL STATUS

[Date of request for examination]

20.01.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

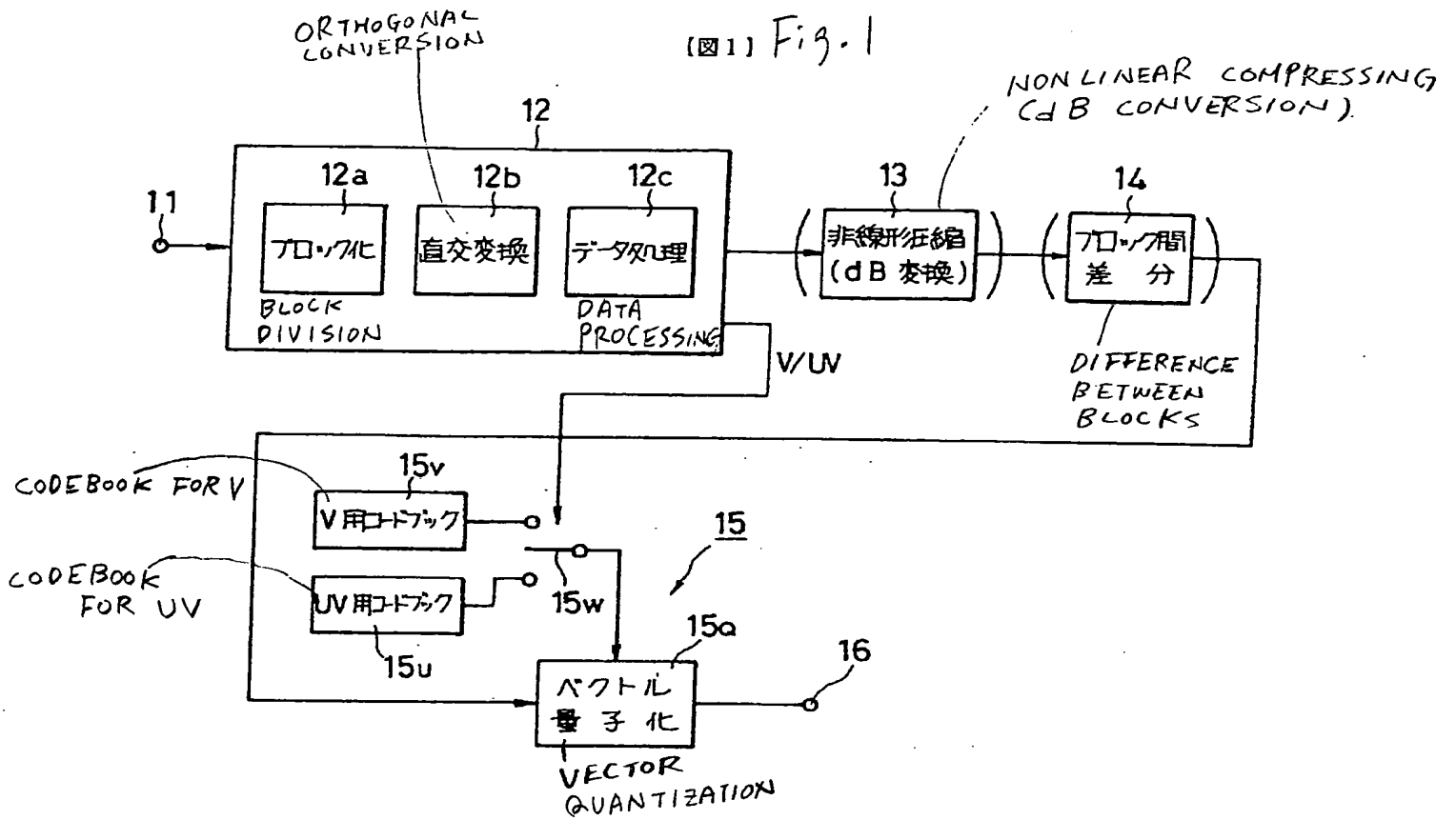
[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

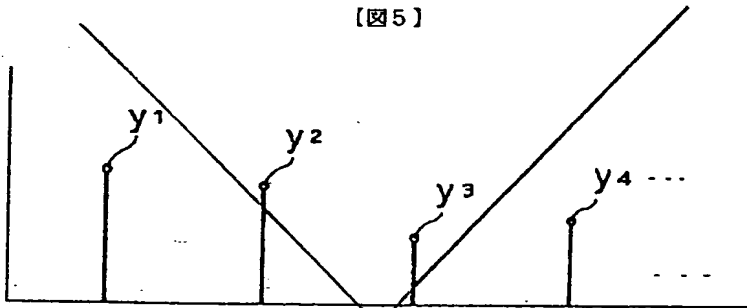
[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

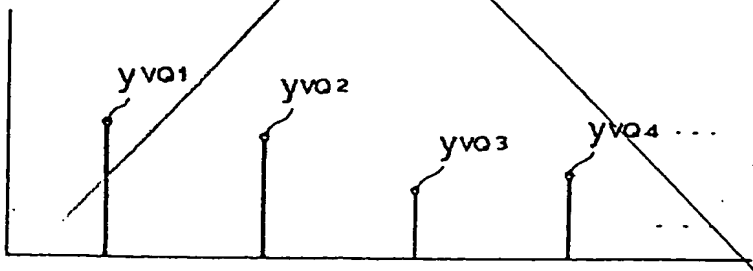
【図1】 Fig. 1



【図5】



【図6】



Partial Translation of HEI 5-265499

[0009] The above-stated problem is solved by an efficient coding method including a step of obtaining data on a frequency axis as M-dimensional vector based on data obtained by dividing an input audio signal into block units and converting to the frequency axis, a step of lowering a dimension to S-dimension ($S < M$) by obtaining a representative value for each of groups by dividing the data on the frequency axis of the M-dimensional vector into a plurality of groups, a step of performing first vector quantization for data of S-dimensional vector, a step of obtaining corresponding S-dimensional code vector by dequantizing output data from first vector quantization, a step of extending the S-dimensional code vector to the original M-dimensional vector, and a step of performing quantization by switching a plurality of code books based on a parameter indicating characteristics of each block of the audio input signal by using a second vector quantization device including the plurality of code books based on a state of an audio signal for performing second vector quantization for data showing relationship between the extended M-dimensional vector and the data on the frequency axis of the M-dimensional data.

[0010] In these inventions, when a voice signal is used as the audio signal, it is possible to use a plurality of code books as the code book depending if the voice signal is voiced sound or unvoiced sound and use a parameter indicating if an input voice signal in the each block is the voiced sound or unvoiced sound . Further, a pitch value, strength of pitch component, inclusion ratio of voiced sound / unvoiced sound, gratitude or level of signal spectrum, etc. can be used as a characteristics parameter. Basically, it is better to switch the code books depending if the voice signal is voice sound or unvoiced sound. It is possible to

transmit the characteristics parameter independently. However, it is also possible to substitute a parameter which is predefined in the coding method and transmitted anyway. Further, it is also possible to use data produced by converting to the frequency axis in the block unit and performing non-linear compressing as data on the frequency axis of the M-dimensional vector. Further, before the vector quantization, it is possible to get difference between blocks of the data of which vector quantization is to be performed and perform the vector quantization for the data of the difference between blocks.

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)特許出願公開番号

特開平5-265499

(43)公開日 平成5年(1993)10月15日

(51)Int.Cl.⁵

G10L 9/18
9/00

識別記号

庁内整理番号

FI

技術表示箇所

E 8946-5H
G 8946-5H
J 8946-5H

審査請求 未請求 請求項の数3(全30頁)

(21)出願番号 特願平4-92258

(22)出願日 平成4年(1992)3月18日

(71)出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72)発明者 西口 正之

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(72)発明者 松本 淳

東京都品川区北品川6丁目7番35号 ソニー株式会社内

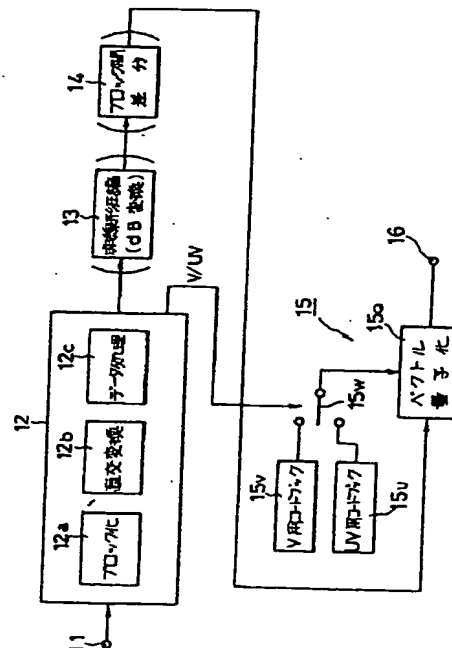
(74)代理人 弁理士 小池 晃 (外2名)

(54)【発明の名称】 高能率符号化方法

(57)【要約】

【構成】 入力された音声信号を周波数変換処理部12で周波数軸上データに変換し、ベクトル量子化部15でベクトル量子化を施して出力端子16から取り出す。ベクトル量子化部15は、V(有声音)用コードブック15_vと、UV(無声音)用コードブック15_uとが切換スイッチ15_{sw}にて切換選択されるようになっており、周波数変換処理部12からのV(有声音)/UV(無声音)判別情報に応じて切換スイッチ15_{sw}が切換制御される。

【効果】 ベクトル量子化の品質が向上し、性能向上に役立つ。



【特許請求の範囲】

【請求項1】 入力されたオーディオ信号をブロック単位で区分して周波数軸に変換して得られるデータに基づいてM次元ベクトルとしての周波数軸上データを求める工程と、

このM次元ベクトルの周波数軸上データに対してベクトル量子化を施すためにオーディオ信号の状態に応じた複数のコードブックを有するベクトル量子化器を用い、上記オーディオ入力信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切り換えて量子化を施す工程とを有することを特徴とする高能率符号化方法。

【請求項2】 上記オーディオ信号として音声信号を用い、上記コードブックとして音声信号が有声音か無声音かに応じた複数のコードブックを用い、上記特徴パラメータとして上記ブロック毎の入力音声信号が有声音か無声音かを表すパラメータを用いることを特徴とする請求項1記載の高能率符号化方法。

【請求項3】 入力されたオーディオ信号をブロック単位で区分して周波数軸に変換して得られるデータに基づいてM次元ベクトルとしての周波数軸上データを求める工程と、

このM次元ベクトルの周波数軸上データを複数グループに分割して各グループ毎に代表値を求めることにより次元をS次元 ($S < M$) に低下させる工程と、

このS次元ベクトルのデータに対して第1のベクトル量子化を施す工程と、

この第1のベクトル量子化出力データを逆量子化して対応するS次元のコードベクトルを求める工程と、

このS次元のコードベクトルを元のM次元のベクトルに拡張する工程と、

この拡張されたM次元のベクトルと元の上記M次元ベクトルの周波数軸上データとの関係を表すデータに対して第2のベクトル量子化を施すためにオーディオ信号の状態に応じた複数のコードブックを有する第2のベクトル量子化器を用い、上記オーディオ入力信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切り換えて量子化を施す工程とを有することを特徴とする高能率符号化方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、入力された音声信号や音響信号等のオーディオ信号をフレーム単位で区分して周波数軸上のデータに変換して符号化を施すような高能率符号化方法に関する。

【0002】

【従来の技術】 オーディオ信号（音声信号や音響信号を含む）の時間領域や周波数領域における統計的性質と人間の聴感上の特性を利用して信号圧縮を行うような符号化方法が種々知られている。この符号化方法としては、

大別して時間領域での符号化、周波数領域での符号化、分析合成符号化等が挙げられる。

【0003】 音声信号等の高能率符号化の例として、MBE (Multiband Excitation: マルチバンド励起) 符号化、SBE (Singleband Excitation: シングルバンド励起) 符号化、ハーモニック (Harmonic) 符号化、SBC (Sub-band Coding: 帯域分割符号化)、LPC (Linear Predictive Coding: 線形予測符号化)、あるいはDCT (離散コサイン変換)、MDCT (モデファイドDCT)、FFT (高速フーリエ変換) 等において、スペクトル振幅やそのパラメータ (LSPパラメータ、 α パラメータ、 k パラメータ等) のような各種情報データを量子化する場合に、従来においてはスカラ量子化を行うことが多い。

【0004】

【発明が解決しようとする課題】 ところで、ビットレートを例えば3~4 kbps程度にまで低減し、量子化効率を更に向上させようとする、スカラ量子化では量子化雑音（歪み）が大きくなってしまい、実用化が困難であった。そこで、これらの符号化の際に得られる時間軸データや周波数軸データやフィルタ係数データ等を個々に量子化せず、複数のデータの組（ベクトル）にまとめて一つの符号で表現して量子化するベクトル量子化が注目されている。

【0005】 ここで、上記MBE、SBE、LPC等のスペクトルエンベロープ、あるいはそのパラメータ (LSPパラメータ、 α パラメータ、 k パラメータ等) のベクトル量子化の際には、固定のコードブックを使用している。しかしながら、使用可能なビット数が減少（低ビットレート化）してくると、固定コードブックでは十分な性能が得られなくなる。このため、ベクトル量子化する入力データのベクトル空間上での存在領域が狭くなるようにクラスタリング（分類分け）したものをベクトル量子化することが好ましい。

【0006】 また、伝送ビットレートに余裕があるときも、コードブックサイズ、サーチのための演算量を低減するため、構造化されたコードブックを使用することも考えられているが、このとき、出力インデックス長が $n+1$ ビットの1個のコードブックを使用する代わりに、例えば2個の n ビットのコードブックに分割する方が好ましい。

【0007】 本発明は、このような実情に鑑みてなされたものであり、ベクトル量子化を入力データの性質等に応じて効率良く行え、ベクトル量子化器のコードブックのサイズやサーチ時の演算量を低減でき、品質の高い符号化が行えるような高能率符号化方法の提供を目的とする。

【0008】

【課題を解決するための手段】 本発明に係る高能率符号化方法は、入力されたオーディオ信号（音声信号や音響

信号等)をブロック単位で区分して周波数軸に変換して得られるデータに基づいてM次元ベクトルとしての周波数軸上データを求める工程と、このM次元ベクトルの周波数軸上データに対してベクトル量子化を施すためにオーディオ信号の状態に応じた複数のコードブックを有するベクトル量子化器を用い、上記オーディオ入力信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切り換えて量子化を施す工程とを有することにより、上述の課題を解決するものである。

【0009】他の発明に係る高能率符号化方法としては、入力されたオーディオ信号をブロック単位で区分して周波数軸に変換して得られるデータに基づいてM次元ベクトルとしての周波数軸上データを求める工程と、このM次元ベクトルの周波数軸上データを複数グループに分割して各グループ毎に代表値を求めることにより次元をS次元($S < M$)に低下させる工程と、このS次元ベクトルのデータに対して第1のベクトル量子化を施す工程と、この第1のベクトル量子化出力データを逆量子化して対応するS次元のコードベクトルを求める工程と、このS次元のコードベクトルを元のM次元のベクトルに拡張する工程と、この拡張されたM次元のベクトルと元の上記M次元ベクトルの周波数軸上データとの関係を表すデータに対して第2のベクトル量子化を施すためにオーディオ信号の状態に応じた複数のコードブックを有する第2のベクトル量子化器を用い、上記オーディオ入力信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切り換えて量子化を施す工程とを有することにより、上述の課題を解決する。

【0010】これらの発明において、上記オーディオ信号として音声信号を用る場合に、上記コードブックとして音声信号が有声音か無声音かに応じた複数のコードブックを用い、上記特徴パラメータとして上記ブロック毎の入力音声信号が有声音か無声音かを表すパラメータを用いることができる。また、特徴パラメータとして、ピッチの値、ピッチの成分の強弱、有声音/無声音の含有比率、信号スペクトルの傾き及びそのレベル等が使用でき、基本的には有声音か無声音かに応じてコードブックを切り換えることが好ましい。このような特徴パラメータは、別途伝送してもよいが、符号化方式により予め規定されているような元々伝送されるパラメータを代用させるようにしてもよい。また、上記M次元ベクトルの周波数軸上データとして、上記ブロック単位で周波数軸に変換したデータを非線形圧縮したものを用いることができる。さらに、上記ベクトル量子化の前に、ベクトル量子化しようとするデータのブロック間の差分をとり、このブロック間差分データに対してベクトル量子化を施すようにしてもよい。

【0011】

【作用】入力されたオーディオ信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切

り換えてベクトル量子化を施すことにより、量子化が効率良く行え、ベクトル量子化器のコードブックのサイズやサーチ時の演算量を低減でき、品質の高い符号化が行える。

【0012】

【実施例】以下、本発明に係る高能率符号化方法の実施例について、図面を参照しながら説明する。図1は本発明の一実施例となる高能率符号化方法を説明するための符号化装置(エンコーダ)の概略構成を示している。

【0013】図1において、入力端子11には、音声信号あるいは音響信号が供給され、この入力信号は、周波数軸変換処理部12にて周波数軸上のスペクトル振幅データに変換される。この周波数軸変換処理部12の内部には、例えば入力された時間軸上の信号の所定数のサンプル(Nサンプル)毎にブロック化するブロック化部12a、FFT(高速フーリエ変換)等の直交変換部12b、スペクトルエンベロープの特徴を表す振幅情報を求めるためのデータ処理部12c等が設けられている。周波数軸変換処理部12からの出力は、必要に応じて例えばdB領域に変換する非線形圧縮部13を介し、必要に応じてブロック間の差分をとる処理部14を介してベクトル量子化部15に送られる。このベクトル量子化部15では、入力データの所定数(Mサンプル)がまとめられてM次元とベクトルとされ、ベクトル量子化処理が施される。このようなM次元ベクトル量子化処理は、一般的には、入力された次元ベクトルに対してM次元空間上で最も距離が近いコードベクトルがコードブックから検索(サーチ)されて、該検索されたコードベクトルのインデックスを出力端子16から取り出すような処理であるが、この図1に示す実施例のベクトル量子化部15は、複数種類のコードブックを有し、これらのコードブックが、周波数軸変換処理部12からの上記入力信号の特徴を表すパラメータに応じて切り換えられるようになっている。

【0014】図1の例では、入力信号として音声信号を想定し、V(有声音)用コードブック15、と、UV(無声音)用コードブック15、とが切換スイッチ15により切り換えられてベクトル量子化器15に送られるようになっており、切換スイッチ15は、周波数軸変換処理部12からのV/UV(有声音/無声音)判別信号に応じて切換制御される。このV/UV判別信号(フラグ等)は、後述するようなMBE(Multiband Excitation: マルチバンド励起)ポコード(音声分析合成装置)等の場合には、分析系(エンコーダ)側から合成系(デコーダ)側へ伝送されるパラメータであり、別途に伝送する必要がない。

【0015】ここで上記MBEの場合を例にとって説明すると、上記コードブック15、15の切換のためのパラメータは、伝送される情報の一つであるV/UV判別フラグを利用するようにすればよい。すなわち、上

記周波数軸変換処理部12において、ピッチに応じたバンド分割を行い、分割された各バンド毎にV（有声音）かUV（無声音）かの判別が行われる。ここで、Vバンドの個数を N_v 、UVバンドの個数を N_{uv} とし、所定の*

$$\frac{N_v}{N_v + N_{uv}} \geq V_{th} \quad \dots (1)$$

のときは、V（有声音）用のコードブック15_vを切換選択するようにし、それ以外のときはUV（無声音）用コードブック15_uを切換選択するようにする。上記閾値 V_{th} は、例えば1程度に設定すればよい。

【0017】デコーダ（合成系）側でも同様に、V（有声音）用及びUV（無声音）用の2種類のコードブックの切換選択を行う。上記V/UV判別フラグはMBEポコーダでは必ず伝送するサイドインフォメーション（補助情報）であるので、この具体例においては別個にコードブック切換のための特徴パラメータを伝送する必要がなく、伝送ビットレートの増加がない。

【0018】V（有声音）用コードブック15_v、UV（無声音）用コードブック15_uの生成（トレーニング）は、単に同じ基準でトレーニングデータを分けることによって可能となる。すなわち、V（有声音）と判別された振幅データの集まりから生成されたコードブック※

$$P_L = \sum_{i=0}^{(N/4)-1} \text{rms}^2(i)$$

$$P_H = \sum_{i=N/4}^{(N/2)-1} \text{rms}^2(i)$$

ただし、 $\text{rms}(i)$ は、 $\sqrt{\text{Re}^2(i) + \text{Im}^2(i)}$

$\text{Re}(i)$ 及び $\text{Im}(i)$ は入力数列のFFTの実部及び虚部である。

この(2)式の P_L 、 P_H を用いて、
【0022】

$$R_d = \frac{P_L}{P_H}$$

$$L = \sqrt{\frac{P_L + P_H}{N/2}}$$

として、 $R_d < R_{th}$ 、かつ $L < L_{th}$ のときは、無条件に全バンドUV（無声音）と判別する。

【0023】これを行うと、間違ったピッチを微小入力で検出したときに、それを使用しない効果がある。この

* 閾値 V_{th} に対して、

【0016】

【数1】

※をV用コードブック15_vとし、UV（無声音）と判別された振幅データの集まりから生成されたコードブックをUV用コードブック15_uとする。

【0019】この具体例においては、V/UV情報をコードブックの切換に用いていることから、上記V/UV判別フラグをより確実に（信頼性を高く）しておく必要がある。例えば、明らかに子音や背景雑音であるとみなせる部分では、全バンドUVにすべきである。その判断の一例としては、高域にパワーの多い微小入力をUV化することが挙げられる。

【0020】入力信号のN点（256サンプル）に対してFFT（高速フーリエ変換）を行い、有効な $0 \sim \pi$ （ $0 \sim N/2$ ）の中で、 $0 \sim N/4$ と $N/4 \sim N/2$ とのそれぞれの区間でパワー計算を行う。

【0021】

【数2】

... (2)

★【数3】

★

... (3)

ようにして、より確かなV/UVフラグを作っておくと、ベクトル量子化の際のコードブックの切換に好都合である。

【0024】次に、上記V（有声音）用及びUV（無声

音)用の各コードブックを作成する際のトレーニングについて、図2を参照しながら説明する。図2において、トレーニング用の数分程度の音声信号から成るトレーニングセット31からの信号は、周波数軸変換処理部32に送られて、ピッチ抽出部32aによりピッチ抽出が、スペクトル振幅算出部32bによりスペクトル振幅の算出が、またバンド毎のV/U判断部32cにより各バンド毎のV/U判断がそれぞれ行われる。この周波数軸変換処理部32からの出力データは、トレーニング前処理部34に送られる。

【0025】トレーニング前処理部34では、V/U含有率のチェック部34aにより上記(1)式や(4)式の条件等をチェックし、得られたV/U情報に応じて、トレーニングデータの振り分け部34bにより上記スペクトル振幅データを振り分けており、V(有声音)のときにはV(有声音)用トレーニングデータ出力部36aへ、UV(無声音)のときにはUV(有声音)用トレーニングデータ出力部37aへ、それぞれ振幅データを送っている。

【0026】V用トレーニングデータ出力部36aから出力されたV(有声音)のスペクトル振幅データは、トレーニング処理部36bに送られて、例えばいわゆるLBG法によりトレーニング処理が行われ、V(有声音)用コードブック36cが作成される。ここで、上記LBG法とは、Linde, Y., Buzo, A. and Gray, R. M., "An Algorithm for Vector Quantizer Design", IEEE Trans. Comm., COM-28, pp.84-95, Jan. 1980 において提案されたベクトル量子化器を設計するアルゴリズムにおけるコードブックのトレーニング法であり、確率密度関数が知られていない情報源に対していわゆるトレーニング系列を用いて局所的に最適なベクトル量子化器を設計するものである。同様に、UV用トレーニングデータ出力部37aから出力されたUV(無声音)のスペクトル振幅データは、トレーニング処理部37bに送られて例えばLBG法によりトレーニング処理が行われ、UV(無声音)用コードブック37cが作成される。

【0027】ここで、後述するように、ベクトル量子化部に階層構造を持たせ、上位層にはV/U共用部分のコードブックを用い、下位層のコードブックのみをV/Uに応じて切り換えるようにする場合には、V/U共用部分のコードブックも作成する必要がある。この場合、上記周波数軸変換処理部32からの出力データをV/U共用部分のコードブック用トレーニングデータ出力部35aに送ることが必要とされる。

【0028】V/U共用部分のコードブック用トレーニングデータ出力部35aから出力されたスペクトル振幅データは、トレーニング処理部35bに送られて、例えばいわゆるLBG法によりトレーニング処理が実行され、V/U共用部分のコードブック35cが作成される。この作成されたV/U共用コードブック35cか

らのコードベクトルを、上記V用、UV用の各トレーニングデータ出力部36a、37aに送り、入力データから減算することが必要とされる。

【0029】以下、階層構造化されたベクトル量子化部の具体的構成及び作用について、図3～図9を参照しながら説明する。すなわち、この図3に示すベクトル量子化部15は、例えば上下の2つの層に階層構造化されており、入力ベクトルに対して2段階のベクトル量子化が施されるようになっている。

10 【0030】図3に示すベクトル量子化器15の入力端子17には、上記図1の周波数軸変換処理部12からの周波数軸上の振幅データが(必要に応じて非線形圧縮部13及びブロック間差分処理部14を介し)、ベクトル量子化の単位となる上記M次元ベクトルとして供給されている。このM次元ベクトルは、次元低減部21に送られることにより複数グループに分割され、各グループ毎に代表値を求めることにより次元がS次元($S < M$)に低下せられる。ここで図4は、ベクトル量子化器15に入力されるM次元ベクトルの各要素、すなわち周波数軸上のM個の振幅データ $x(n)$ の具体例を示しており、 $1 \leq n \leq M$ である。これらのM個の振幅データ $x(n)$ は、例えば4サンプル毎にまとめられてそれぞれの代表値、例えば平均値 y_i が求められ、図5に示すように、平均値データ y_i 、 $\sim y_s$ のS個(この場合は $S = M/4$)から成るS次元ベクトルが得られる。

30 【0031】次に、このS次元ベクトルのデータに対してS次元ベクトル量子化器22によりベクトル量子化が施される。すなわち、S次元ベクトル量子化器22のコードブック22。中のS次元コードベクトルの内、入力されたS次元ベクトルにS次元空間上で最も距離が近いコードベクトルがサーチされ、このサーチされたコードベクトルのインデックスデータは出力端子26より取り出され、サーチされたコードベクトル(出力インデックスを逆ベクトル量子化して得られるコードベクトル)は次元拡張部23に送られる。コードブック22。には、上記図2において説明したV/U共用部分のコードブック35cが用いられる。図6は、上記図5に示すS個の平均値データ y_i 、 $\sim y_s$ から成るS次元ベクトルをベクトル量子化した後、逆量子化して(あるいはベクトル量子化器22のコードブックで上記量子化の際にサーチされたコードベクトルを取り出して)得られたローカルデコード出力としてのS次元ベクトル v_q の各要素 $y_{vq1} \sim y_{vqs}$ を示している。

40 【0032】次に次元拡張部23では、上記S次元のコードベクトルを元のM次元のベクトルに拡張する。この拡張されたM次元ベクトルの各要素の例を図7に示す。この図7から明らかなように、上記逆ベクトル量子化されたS次元ベクトル v_q の各要素 $y_{vq1} \sim y_{vqs}$ をそれぞれ元の4サンプルずつに増加させることにより、4S=M個の要素から成るM次元ベクトルを得るわけであ

る。この拡張されたM次元のベクトルと元の上記M次元ベクトルの周波数軸上データとの関係を表すデータに対して第2のベクトル量子化を施す。

【0033】図3の具体例では、次元拡張部23からの拡張されたM次元ベクトルデータを減算器24に送り、元の上記M次元ベクトルの周波数軸上データから減算することにより、上記S次元を拡張したM次元ベクトルと元のM次元ベクトルとの関係を表すS個のベクトルデータを得ている。図8は、上記図4に示すM次元ベクトル

の各要素である周波数軸上のM個の振幅データ x (n) から、図7に示す拡張M次元ベクトルの各要素を減算して得られたM個のデータ $r_1 \sim r_M$ を示しており、これらのM個のデータ $r_1 \sim r_M$ の4サンプルずつを組(ベクトル)としてS個の4次元ベクトル $v_{q1} \sim v_{qs}$ が得られる。

【0034】このようにして減算器24から得られたS個のベクトルのそれぞれに対して、ベクトル量子化器群25のS個の各ベクトル量子化器25_{1q}~25_{sq}によりそれぞれベクトル量子化を施す。各ベクトル量子化器25_{1q}~25_{sq}から出力されるインデックスは、それぞれ出力端子27_{1q}~27_{sq}から取り出される。図9は、ベクトル量子化器25_{1q}~25_{sq}としてそれぞれ4次元ベクトル量子化器を用い、上記図8に示す各4次元ベクトル $v_{q1} \sim v_{qs}$ をそれぞれベクトル量子化した後の各4次元ベクトル $v_{q1} \sim v_{qs}$ の各要素 $r_{vq1} \sim r_{vqs}$ を示している。

【0035】これらの各ベクトル量子化器25_{1q}~25_{sq}には、それぞれV(有声音)用のコードブック25_{1v}~25_{sv}とUV(無声音)用のコードブック25_{1u}~25_{su}とが用いられ、これらのV用コードブック25_{1v}~25_{sv}とUV用コードブック25_{1u}~25_{su}とは、入力端子18からのV/UV情報に応じて切換制御される切換スイッチ25_{1s}~25_{ss}により切換選択されるようになっている。これらの切換スイッチ25_{1s}~25_{ss}の切換制御は、全バンドに対して同時に(連動させて)行ってもよいが、各ベクトル量子化器25_{1q}~25_{sq}が受け持つ周波数帯域が異なることを考慮して、バンド毎のV/UV判別フラグに応じて切換制御するようにしてもよい。V用コードブック25_{1v}~25_{sv}が図2のV(有声音)用コードブック36cに対応し、UV用コードブック25_{1u}~25_{su}が図2のV(有声音)用コードブック37cに対応することは勿論である。

【0036】このような階層構造化された2段階のベクトル量子化を施すことにより、コードブックサーチのための演算量を低減でき、コードブックのためのメモリ量(例えばROM容量)を低減でき、また上記出力端子26から得られる上位層のより重要なインデックスに対して誤り訂正符号化を施して重点的に保護するようにすること等により、誤り訂正符号の効果的な適用が可能となる。なお、ベクトル量子化部15の階層構造は、2段階

に限定されず、3段階以上の多層の階層構造を持たせるようにしてもよい。

【0037】なお、図1~図3の各部は、全てをハードウェアにて構成する必要はなく、例えばいわゆるDSP(ディジタル信号プロセッサ)等を用いてソフトウェア的に実現するようにしてもよい。

【0038】以上説明したように、例えば音声合成分析符号化の場合に、音声の有声無声の度合い、ピッチ等が既に特徴量として抽出されていることを考慮して、それらの特徴量、特に有聲/無声の判断結果によってベクトル量子化のコードブックを切り換えることにより、良好なベクトル量子化が実現できる。すなわち、有声音時と無声音時とは、各々存在するスペクトルの形状が大きく異なり、各状態に対応して各々別々にトレーニングされたコードブックを持つことは、特性改善上、大変好ましいことである。また、階層構造化ベクトル量子化の場合、上位層のベクトル量子化は固定のコードブックとし、下位層のベクトル量子化のコードブックのみ、有声音/無声音用の2種を切り換えるようにしてもよい。また、周波数軸上のビットアロケーションも、例えば有声音のときは低音重視、無声音のときは高音重視に切り換えるようにしてもよい。切換制御には、ピッチの有無、有声音/無声音の比率、スペクトルのレベルや傾き、等を利用することができる。さらに、3種類以上のコードブックを切り換えるようにしてもよく、例えば、無声音用のコードブックを子音、背景雑音等で2種類以上用いるようにしてもよい。

【0039】次に、上述したような高能率符号化方法が適用可能な、音声信号の合成分析符号化装置(いわゆるボコーダ)の一種のMBE(Multiband Excitation: マルチバンド励起)ボコーダの具体例について、図面を参照しながら説明する。以下に説明するMBEボコーダは、D.W. Griffin and J.S. Lim, "Multiband Excitation Vocoder," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36, No. 8, pp. 1223-1235, Aug. 1988に開示されているものであり、従来のPARCOR(Partial auto-CORrelation: 偏自己相関)ボコーダ等では、音声のモデル化の際に有声音区間と無声音区間とをブロックあるいはフレーム毎に切り換えていたのに対し、MBEボコーダでは、同時刻(同じブロックあるいはフレーム内)の周波数軸領域に有声音(Voiced)区間と無声音(Unvoiced)区間とが存在するという仮定でモデル化している。

【0040】図10は、上記MBEボコーダに本発明を適用した実施例の全体の概略構成を示すブロック図である。この図10において、入力端子101には音声信号が供給されるようになっており、この入力音声信号は、HPF(ハイパスフィルタ)等のフィルタ102に送られて、いわゆるDC(直流)オフセット分の除去や帯域制限(例えば200~3400Hzに制限)のための少な

11

くとも低域成分(200Hz以下)の除去が行われる。このフィルタ102を介して得られた信号は、ピッチ抽出部103及び窓かけ処理部104にそれぞれ送られる。ピッチ抽出部103では、入力音声信号データが所定サンプル数N(例えばN=256)単位でブロック分割され(あるいは方形窓による切り出しが行われ)、このブロック内の音声信号についてのピッチ抽出が行われる。このような切り出しブロック(256サンプル)を、例えば図11のAに示すようにLサンプル(例えばL=1*

$$x_s(k, q) = x(q) w(kL - q)$$

となる。この(4)式において、kはブロック番号、qはデータの時間インデックス(サンプル番号)を表し、処理前の入力信号のq番目のデータx(q)に対して第kブロックの窓(ウィンドウ)関数w(kL-q)により窓*

$$w_r(r) = \begin{cases} 1 & 0 \leq r < N \\ 0 & r < 0, N \leq r \end{cases} \quad \dots (5)$$

また、窓かけ処理部104での図11のBに示すような★ハミング窓の場合の窓関数w_r(r)は、

$$w_h(r) = \begin{cases} 0.54 - 0.46 \cos(2\pi r/(N-1)) & 0 \leq r < N \\ 0 & r < 0, N \leq r \end{cases} \quad \dots (6)$$

である。このような窓関数w_r(r)あるいはw_h(r)を用いるときの上記(4)式の窓関数w(r)(=w(kL-q))の非零区間は、

$$0 \leq kL - q < N$$

これを变形して、

$$kL - N < q \leq kL$$

従って例えば上記方形窓の場合に窓関数w_r(kL-q)=1となるのは、図12に示すように、kL-N < q ≤ kLのときとなる。また、上記(4)～(6)式は、長さN(=256)サンプルの窓が、L(=160)サンプルずつ前進してゆくことを示している。以下、上記(5)式、(6)式の各窓関数で切り出された各N点(0 ≤ r < N)の非零サンプル列を、それぞれx_{s,r}(k, r)、x_{h,r}(k, r)と表すことにする。

【0042】窓かけ処理部104では、図13に示すように、上記(6)式のハミング窓がかけられた1ブロック256サンプルのサンプル列x_{h,r}(k, r)に対して1792サンプル分の0データが付加されて(いわゆる0詰めされて)2048サンプルとされ、この2048サンプルの時間軸データ列に対して、直交変換部105により例えばFFT(高速フーリエ変換)等の直交変換処理が施される。

【0043】ピッチ抽出部103では、上記x_{s,r}(k, r)のサンプル列(1ブロックNサンプル)に基づいてピッチ抽出が行われる。このピッチ抽出法には、時間波形の周期性や、スペクトルの周期的周波数構造や、自己相関関数を用いるもの等が知られているが、本実施例では、センタクリップ波形の自己相関法を採用している。このときのブロック内でのセンタクリップレベルについては、1ブロックにつき1つのクリップレベルを設定して

12

*60)のフレーム間隔で時間軸方向に移動させており、各ブロック間のオーバーラップはN-Lサンプル(例えば96サンプル)となっている。また、窓かけ処理部104では、1ブロックNサンプルに対して所定の窓関数、例えばハミング窓をかけ、この窓かけブロックを1フレームLサンプルの間隔で時間軸方向に順次移動させている。

【0041】このような窓かけ処理を数式で表すと、

$$\dots (4)$$

窓かけ処理されることによりデータx_s(k, q)が得られることを示している。ピッチ抽出部103内での図11のAに示すような方形窓の場合の窓関数w_r(r)は、

$$\dots (5)$$

$$\dots (6)$$

もよいが、ブロックを細分割した各部(各サブブロック)の信号のピークレベル等を検出し、これらの各サブブロックのピークレベル等の差が大きいために、ブロック内でクリップレベルを段階的にあるいは連続的に変化させるようにしている。このセンタクリップ波形の自己相関データのピーク位置に基づいてピーク周期を決めている。このとき、現在フレームに属する自己相関データ(自己相関は1ブロックNサンプルのデータを対象として求められる)から複数のピークを求めておき、これらの複数のピークの内の最大ピークが所定の閾値以上のときには該最大ピーク位置をピッチ周期とし、それ以外のときには、現在フレーム以外のフレーム、例えば前後のフレームで求められたピッチに対して所定の関係を満たすピッチ範囲内、例えば前フレームのピッチを中心として±20%の範囲内にあるピークを求め、このピーク位置に基づいて現在フレームのピッチを決定するようにしている。このピッチ抽出部103ではオープンループによる比較的ラフなピッチのサーチが行われ、抽出されたピッチデータは高精度(ファイン)ピッチサーチ部106に送られて、クローズドループによる高精度のピッチサーチ(ピッチのファインサーチ)が行われる。

【0044】高精度(ファイン)ピッチサーチ部106には、ピッチ抽出部103で抽出された整数(インテジャ)値の粗(ラフ)ピッチデータと、直交変換部105により例えばFFTされた周波数軸上のデータとが供給されている。この高精度ピッチサーチ部106では、上記粗ピッチデータ値を中心に、0.2～0.5きざみで±数サンプルずつ振って、最適な小数点付き(フローティング)のファインピッチデータの値へ追い込む。このときのファインサーチの手法として、いわゆる合成による

分析 (Analysis by Synthesis)法を用い、合成されたパワースペクトルが原音のパワースペクトルに最も近くなるようにピッチを選んでいく。

【0045】このピッチのファインサーチについて説明*

$$S(j) = H(j) | E(j) |$$

と表現するようなモデルを想定している。ここで、 J は $\pi \omega_s = f_s / 2$ に対応し、サンプリング周波数 $f_s = 2 \pi \omega_s$ が例えば 8 kHz のときには 4 kHz に対応する。上記 (7) 式中において、周波数軸上のスペクトルデータ $S(j)$ が図 14 の A に示すような波形のとき、 $H(j)$ は、図 14 の B に示すような元のスペクトルデータ $S(j)$ のスペクトル包絡線 (エンベロープ) を示し、 $E(j)$ は、図 14 の C に示すような等レベルで周期的な励起信号 (エキサイテーション) のスペクトルを示している。すなわち、FFT スペクトル $S(j)$ は、スペクトルエンベロープ $H(j)$ と励起信号のパワースペクトル $|E(j)|$ との積としてモデル化される。

【0046】上記励起信号のパワースペクトル $|E(j)|$ は、上記ピッチに応じて決定される周波数軸上の波形の周期性 (ピッチ構造) を考慮して、1つの帯域 (バンド) の波形に相当するスペクトル波形を周波数軸上の各*

$$\varepsilon_m = \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \} \quad \dots (8)$$

で表せる。このエラー ε_m を最小化するような $|A_m|$ は、

$$\frac{\partial \varepsilon_m}{\partial |A_m|} = 2 \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \} |E(j)|$$

$$\therefore |A_m| = 2 \sum_{j=a_m}^{b_m} |S(j)| |E(j)| / \sum_{j=a_m}^{b_m} |E(j)|^2 \quad \dots (9)$$

となり、この (6) 式の $|A_m|$ のとき、エラー ε_m を最小化する。このような振幅 $|A_m|$ を各バンド毎に求め、得られた各振幅 $|A_m|$ を用いて上記 (8) 式で定義された各バンド毎のエラー ε_m を求める。次に、このような各バンド毎のエラー ε_m の全バンドの総和値 $\Sigma \varepsilon$ を求める。さらに、このような全バンドのエラー総和値 $\Sigma \varepsilon$ を、いくつかの微小に異なるピッチについて求め、エラー総和値 $\Sigma \varepsilon$ が最小となるようなピッチを求める。

【0050】すなわち、上記ピッチ抽出部 103 で求められたラフピッチを中心として、例えば 0.25 きざみで上下に数種類ずつ用意する。これらの複数種類の微小に異なるピッチの各ピッチに対してそれぞれ上記エラー総和値 $\Sigma \varepsilon$ を求める。この場合、ピッチが定まるとバンド幅が決まり、上記 (9) 式より、周波数軸上データのパワースペクトル $|S(j)|$ と励起信号スペクトル $|E(j)|$ とを用いて上記 (8) 式のエラー ε_m を求め、そ

*する。まず、上記 MBE ボコーダにおいては、上記 FFT 等により直交変換された周波数軸上のスペクトルデータとしての $S(j)$ を

$$0 < j < J \quad \dots (7)$$

※バンド毎に繰り返すように配列することにより形成される。この 1 バンド分の波形は、例えば上記図 13 に示すような 256 サンプルのハミング窓関数に 1792 サンプル分の 0 データを付加 (0 詰め) した波形を時間軸信号と見なして FFT し、得られた周波数軸上のある帯域幅を持つインパルス波形を上記ピッチに応じて切り出すことにより形成することができる。

【0047】次に、上記ピッチに応じて分割された各バンド毎に、上記 $H(j)$ を代表させるような (各バンド毎のエラーを最小化するような) 値 (一種の振幅) $|A_m|$ を求める。ここで、例えば第 m バンド (第 m 高調波の帯域) の下限、上限の点をそれぞれ a_m 、 b_m とするとき、この第 m バンドのエラー ε_m は、

【0048】
【数 4】

★【0049】
★【数 5】

の全バンドの総和値 $\Sigma \varepsilon$ を求めることができる。このエラー総和値 $\Sigma \varepsilon$ を各ピッチ毎に求め、最小となるエラー総和値に対応するピッチを最適のピッチとして決定するわけである。以上のようにして高精度ピッチサーチ部 106 で最適のファイン (例えば 0.25 きざみ) ピッチが求められ、この最適ピッチに対応する振幅 $|A_m|$ が決定される。

【0051】以上ピッチのファインサーチの説明においては、説明を簡略化するために、全バンドが有声音 (Voiced) の場合を想定しているが、上述したように MBE ボコーダにおいては、同時刻の周波数軸上に無声音 (Unvoiced) 領域が存在するというモデルを採用していることから、上記各バンド毎に有声音/無声音の判別を行うことが必要とされる。

【0052】上記高精度ピッチサーチ部 106 からの最適ピッチ及び振幅 $|A_m|$ のデータは、有声音/無声音判別部 107 に送られ、上記各バンド毎に有声音/無声

音の判別が行われる。この判別のために、NSR（ノイズとシグナル比）を利用する。すなわち、第 m バンドのNSRは、

$$NSR = \frac{\sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \}^2}{\sum_{j=a_m}^{b_m} |S(j)|^2} \quad \dots (10)$$

と表せ、このNSR値が所定の閾値（例えば0.3）より大のとき（エラーが大きい）ときには、そのバンドでの $|A_m| |E(j)|$ による $|S(j)|$ の近似が良くない（上記励起信号 $|E(j)|$ が基底として不適当である）と判断でき、当該バンドをUV（Unvoiced、無声音）と判別する。これ以外のときは、近似がある程度良好に行われていると判断でき、そのバンドをV（Voiced、有声音）と判別する。

【0054】次に、振幅再評価部108には、直交変換部105からの周波数軸上データ、高精度ピッチサーチ※

$$|A_m|_{vr} = \sqrt{\sum_{j=a_m}^{b_m} |S(j)|^2 / (b_m - a_m + 1)}$$

... (11)

にて求められる。

【0056】この振幅再評価部108からのデータは、データ数変換（一種のサンプリングレート変換）部109に送られる。このデータ数変換部109は、上記ピッチに応じて周波数軸上での分割帯域数が異なり、データ数（特に振幅データの数）が異なることを考慮して、一定の個数にするためのものである。すなわち、例えば有効帯域を3400kHzまでとすると、この有効帯域が上記ピッチに応じて、8バンド～63バンドに分割されることになり、これらの各バンド毎に得られる上記振幅 $|A_m|$ （UVバンドの振幅 $|A_m|_{uv}$ も含む）データの個数 $m_{ux} + 1$ も8～63と変化することになる。このためデータ数変換部109では、この可変個数 $m_{ux} + 1$ の振幅データを一定個数 N_c （例えば44個）のデータに変換している。

【0057】ここで本実施例においては、周波数軸上の有効帯域1ブロック分の振幅データに対して、ブロック内の最後のデータからブロック内の最初のデータまでの値を補間するようなダミーデータを付加してデータ個数を N_c 個に拡大した後、帯域制限型の K_{os} 倍（例えば8倍）のオーバーサンプリングを施すことにより K_{os} 倍の個数の振幅データを求め、この K_{os} 倍の個数（ $(m_{ux} + 1) \times K_{os}$ 個）の振幅データを直線補間してさらに多くの N_c 個（例えば2048個）に拡張し、この N_c 個のデータを間引いて上記一定個数 N_c （例えば44個）のデータに変換する。

*【0053】
【数6】

10※部106からのファインピッチと評価された振幅 $|A_m|$ との各データ、及び上記有声音/無声音判別部107からのV/UV（有声音/無声音）判別データが供給されている。この振幅再評価部108では、有声音/無声音判別部107において無声音（UV）と判別されたバンドに関して、再度振幅を求めている。このUVのバンドについての振幅 $|A_m|_{uv}$ は、
【0055】
【数7】

【0058】このデータ数変換部109からのデータ（上記一定個数 N_c の振幅データ）がベクトル量子化部110に送られて、所定個数のデータ毎にまとめられてベクトルとされ、ベクトル量子化が施される。このベクトル量子化部110として、上述した図1あるいは図3に示すような、V（有声音）用コードブックとUV（無声音）用コードブックとを切換選択するようなベクトル量子化部15を用いている。これらのV用、UV用コードブックは、上記有声音/無声音（V/UV）判別部107からのV/UV判別データに応じて切換制御すればよい。このようなベクトル量子化部110からの量子化出力データは、出力端子111を介して取り出される。

【0059】また、上記高精度のピッチサーチ部106からの高精度（ファイン）ピッチデータは、ピッチ符号化部115で符号化され、出力端子112を介して取り出される。さらに、上記有声音/無声音判別部107からの有声音/無声音（V/UV）判別データは、出力端子113を介して取り出される。これらの各出力端子111～113からのデータは、所定の伝送フォーマットの信号とされて伝送される。

【0060】なお、これらの各データは、上記 N サンプル（例えば256サンプル）のブロック内のデータに対して処理を施すことにより得られるものであるが、ブロックは時間軸上を上記 N サンプルのフレームを単位として前進することから、伝送するデータは上記フレーム単位で得られる。すなわち、上記フレーム周期でピッチデ

ータ、V/UV判別データ、振幅データが更新されることになる。

【0061】次に、伝送されて得られた上記各データに基づき音声信号を合成するための合成側（デコード側）の概略構成について、図15を参照しながら説明する。この図15において、入力端子121には上記ベクトル量子化された振幅データが、入力端子122には上記符号化されたピッチデータが、また入力端子123には上記V/UV判別データがそれぞれ供給される。入力端子121からの量子化振幅データは、逆ベクトル量子化部124に送られて逆量子化される。この逆ベクトル量子化部124も、V（有声音）用コードブックとUV（無声音）用コードブックとを入力端子123から得られるV/UV判別データに応じて切換選択するような構造を有している。逆ベクトル量子化部124からの出力データは、データ数逆変換部125に送られて逆変換され、得られた振幅データが有声音合成部126及び無声音合成部127に送られる。入力端子122からの符号化ピッチデータは、ピッチ復号化部128で復号化され、データ数逆変換部125、有声音合成部126及び無声音合成部127に送られる。また入力端子123からのV/UV判別データは有声音合成部126及び無声音合成部127に送られる。

【0062】有声音合成部126では例えば余弦(cosine)波合成により時間軸上の有声音波形を合成し、無声音*

$$V_s(n) = A_s(n) \cos(\theta_s(n)) \quad 0 \leq n < L \quad \dots (12)$$

と表すことができる。全バンドの内のV（有声音）と判別された全てのバンドの有声音を加算($\Sigma V_s(n)$)して最終的な有声音V(n)を合成する。

【0064】この(12)式中の $A_s(n)$ は、上記合成フレームの先端から終端までの間で補間された第m高調波の振幅である。最も簡単には、フレーム単位で更新され*

$$A_s(n) = (L-n)A_{s0}/L + nA_{sL}/L \quad \dots (13)$$

の式により $A_s(n)$ を計算すればよい。

$$\theta_s(0) = m\omega_{s0}n + n^2 m(\omega_{s1} - \omega_{s0})/2L + \phi_{s0} + \Delta\omega n \quad \dots (14)$$

により求めることができる。この(14)式中で、 ϕ_{s0} は上記合成フレームの先端($n=0$)での第m高調波の位相(フレーム初期位相)を示し、 ω_{s0} は合成フレーム先端($n=0$)での基本角周波数、 ω_{s1} は該合成フレームの終端($n=L$:次の合成フレーム先端)での基本角周波数をそれぞれ示している。上記(11)式中の $\Delta\omega$ は、 $n=L$ における位相 ϕ_{sL} が $\theta_s(L)$ に等しくなるような最小の $\Delta\omega$ を設定する。

【0066】以下、任意の第mバンドにおいて、それぞれ $n=0$ 、 $n=L$ のときのV/UV判別結果に応じた上記振幅 $A_s(n)$ 、位相 $\theta_s(n)$ の求め方を説明する。第mバンドが、 $n=0$ 、 $n=L$ のいずれもV（有声音）とされる場合に、振幅 $A_s(n)$ は、上述した(13)式により、伝送された振幅値 A_{s0} 、 A_{sL} を直線補間して振幅A

*合成部127では例えばホワイトノイズをバンドパスフィルタでフィルタリングして時間軸上の無声音波形を合成し、これらの各有声音合成波形と無声音合成波形とを加算部129で加算合成して、出力端子130より取り出すようにしている。この場合、上記振幅データ、ピッチデータ及びV/UV判別データは、上記分析時の1フレーム(Lサンプル、例えば160サンプル)毎に更新されて与えられるが、フレーム間の連続性を高める(円滑化する)ために、上記振幅データやピッチデータの各値を1フレーム中の例えば中心位置における各データ値とし、次のフレームの中心位置までの間(合成時の1フレーム)の各データ値を補間により求める。すなわち、合成時の1フレーム(例えば上記分析フレームの中心から次の分析フレームの中心まで)において、先端サンプル点での各データ値と終端(次の合成フレームの先端)サンプル点での各データ値とが与えられ、これらのサンプル点間の各データ値を補間により求めるようにしている。

【0063】以下、有声音合成部126における合成処理を詳細に説明する。上記V（有声音）と判別された第mバンド(第m高調波の帯域)における時間軸上の上記1合成フレーム(Lサンプル、例えば160サンプル)分の有声音を $V_s(n)$ とすると、この合成フレーム内の時間インデックス(サンプル番号)nを用いて、

※振幅データの第m高調波の値を直線補間すればよい。すなわち、上記合成フレームの先端($n=0$)での第m高調波の振幅値を A_{s0} 、該合成フレームの終端($n=L$:次の合成フレームの先端)での第m高調波の振幅値を A_{sL} とすると、

$A_s(n)$ を算出すればよい。位相 $\theta_s(n)$ は、 $n=0$ で $\theta_s(0) = \phi_{s0}$ から $n=L$ で $\theta_s(L)$ が ϕ_{sL} となるように $\Delta\omega$ を設定する。

【0067】次に、 $n=0$ のときV（有声音）で、 $n=L$ のときUV（無声音）とされる場合に、振幅 $A_s(n)$ は、 $A_s(0)$ の伝送振幅値 A_{s0} から $A_s(L)$ で0となるように直線補間する。 $n=L$ での伝送振幅値 A_{sL} は無声音の振幅値であり、後述する無声音合成の際に用いられる。位相 $\theta_s(n)$ は、 $\theta_s(0) = \phi_{s0}$ とし、かつ $\Delta\omega = 0$ とする。

【0068】さらに、 $n=0$ のときUV（無声音）で、 $n=L$ のときV（有声音）とされる場合には、振幅 $A_s(n)$ は、 $n=0$ での振幅 $A_s(0)$ を0とし、 $n=L$ で伝送された振幅値 A_{sL} となるように直線補間する。位相 θ

19

20

。(n) については、 $n=0$ での位相 $\theta_0(0)$ として、フ* *レーン終端での位相値 $\phi_{0,0}$ を用いて、

$$\theta_0(0) = \phi_{0,0} - m(\omega_{0,1} + \omega_{1,1})L/2 \quad \dots (15)$$

とし、かつ $\Delta\omega=0$ とする。

※を設定する手法について説明する。上記(14)式で、 $n=L$ と置くことにより、

【0069】上記 $n=0$ 、 $n=L$ のいずれもV(有声音)とされる場合に、 $\theta_0(L)$ が $\phi_{0,0}$ となるように $\Delta\omega$ ※

$$\begin{aligned} \theta_0(L) &= m\omega_{0,1}L + L^2 m(\omega_{1,1} - \omega_{0,1})/2L + \phi_{0,0} + \Delta\omega L \\ &= m(\omega_{0,1} + \omega_{1,1})L/2 + \phi_{0,0} + \Delta\omega L \\ &= \phi_{0,0} \end{aligned}$$

となり、これを整理すると、 $\Delta\omega$ は、

$$\Delta\omega = (\text{mod}2\pi((\phi_{1,0} - \phi_{0,0}) - mL(\omega_{0,1} + \omega_{1,1})/2)/L) \quad \dots (16)$$

となる。この(16)式で $\text{mod}2\pi(x)$ とは、 x の主値を $-\pi \sim +\pi$ の間の値で返す関数である。例えば、 $x=1.3\pi$ のとき $\text{mod}2\pi(x) = -0.7\pi$ 、 $x=2.3\pi$ のとき $\text{mod}2\pi(x) = 0.3\pi$ 、 $x=-1.3\pi$ のとき $\text{mod}2\pi(x) = 0.7\pi$ 、等である。

【0070】ここで、図16のAは、音声信号のスペクトルの一例を示しており、バンド番号(ハーモニクスナンバー) m が8、9、10の各バンドがUV(無声音)とされ、他のバンドはV(有声音)とされている。このV(有声音)のバンドの時間軸信号が上記有声音合成部126により合成され、UV(無声音)のバンドの時間軸信号が無声音合成部127で合成されるわけである。

【0071】以下、無声音合成部127における無声音合成処理を説明する。ホワイトノイズ発生部131からの時間軸上のホワイトノイズ信号波形を、所定の長さ(例えば256サンプル)で適当な窓関数(例えばハミング窓)により窓かけをし、STFT処理部132によりSTFT(ショートタームフーリエ変換)処理を施すことにより、図16のBに示すようなホワイトノイズの周波数軸上のパワースペクトルを得る。このSTFT処理部132からのパワースペクトルをバンド振幅処理部133に送り、図16のCに示すように、上記UV(無声音)とされたバンド(例えば $m=8, 9, 10$)について上記振幅 $|A_{m,n}|$ を乗算し、他のV(有声音)とされたバンドの振幅を0にする。このバンド振幅処理部133には上記振幅データ、ピッチデータ、V/UV判別データが供給されている。バンド振幅処理部133からの出力は、ISTFT処理部134に送られ、位相は元のホワイトノイズの位相を用いて逆STFT処理を施すことにより時間軸上の信号に変換する。ISTFT処理部134からの出力は、オーバーラップ加算部135に送られ、時間軸上で適当な(元の連続的なノイズ波形を復元できるように)重み付けをしながらオーバーラップ及び加算を繰り返す、連続的な時間軸波形を合成する。オーバーラップ加算部135からの出力信号が上記加算部129に送られる。

【0072】このように、各合成部126、127において合成されて時間軸上に戻された有声音部及び無声音部の各信号は、加算部129により適当な固定の混合比

で加算して、出力端子130より再生された音声信号を取り出す。

【0073】なお、上記図10の音声分析側(エンコード側)の構成や図15の音声合成側(デコード側)の構成については、各部をハードウェア的に記載しているが、いわゆるDSP(デジタル信号プロセッサ)等を用いてソフトウェアプログラムにより実現することも可能である。

【0074】なお、本発明は上記実施例のみに限定されるものではなく、例えば、音声信号のみならず、音響信号を入力信号として用いることもできる。また、入力オーディオ信号(音声信号や音響信号)の特徴を表すパラメータとしては、上記V(有声音)/UV(無声音)の判別情報に限定されず、ピッチの値、ピッチ成分の強弱、信号スペクトルの傾きやレベル等を使用することができる。さらに、このような特徴パラメータは、符号化方式に応じて本来伝送するパラメータ情報の一部を代用しても、別途に伝送してもよく、他の伝送パラメータで代用させる場合には適応的コードブックと見なせ、別途伝送する場合には構造化コードブックと見なせる。

【0075】

【発明の効果】以上の説明から明らかなように、本発明に係る高能率符号化方法によれば、ベクトル量子化器に入力されたオーディオ信号(音声信号や音響信号)の状態に応じた複数種類のコードブックを設け、該入力オーディオ信号のブロック毎(フレーム毎)の特徴を表すパラメータに応じて、上記複数種類のコードブックを切り換えてベクトル量子化を施しているため、効率的なベクトル量子化が実現でき、ベクトル量子化の性能向上に役立つ。

【図面の簡単な説明】

【図1】本発明に係る高能率符号化方法が適用される符号化装置(エンコーダ)側の概略構成を示すブロック図である。

【図2】コードブックの形成(トレーニング)方法を説明するための図である。

【図3】本発明の他の実施例の方法を説明するための符号化装置(エンコーダ)の要部の概略構成を示すブロック図である。

【図4】階層構造を有するベクトル量子化の動作を説明するための図である。

【図5】階層構造を有するベクトル量子化の動作を説明するための図である。

【図6】階層構造を有するベクトル量子化の動作を説明するための図である。

【図7】階層構造を有するベクトル量子化の動作を説明するための図である。

【図8】階層構造を有するベクトル量子化の動作を説明するための図である。

【図9】階層構造を有するベクトル量子化の動作を説明するための図である。

【図10】本発明に係る高能率符号化方法が適用される装置の具体例としての音声信号の合成分析符号化装置の分析側（エンコード側）の概略構成を示す機能ブロック図である。

【図11】窓かけ処理を説明するための図である。

【図12】窓かけ処理と窓関数との関係を説明するための図である。

【図13】直交変換（FFT）処理対象としての時間軸データを示す図である。

【図14】周波数軸上のスペクトルデータ、スペクトル包絡線（エンベロープ）及び励起信号のパワースペクトルを示す図である。

【図15】本発明に係る高能率符号化方法が適用される装置の具体例としての音声信号の合成分析符号化装置の合成側（デコード側）の概略構成を示す機能ブロック図である。

*【図16】音声信号を合成する際の無声音合成を説明するための図である。

【符号の説明】

12・・・周波数軸変換処理部

13・・・非線形圧縮部

14・・・ブロック間（フレーム間）差分処理部

15・・・ベクトル量子化部

15_u、25_{1u}～25_{su}・・・ベクトル量子化器

15_v、25_{1v}～25_{sv}・・・V（有声音）用コードブック

10 Dブック

15_u、25_{1u}～25_{su}・・・UV（無声音）用コードブック

15_v、25_{1v}～25_{sv}・・・切換スイッチ

21・・・次元低減部

22_u・・・S次元ベクトル量子化器

22_v・・・S次元コードブック

23・・・次元拡張部

103・・・ピッチ抽出部

104・・・窓かけ処理部

105・・・直交変換（FFT）部

106・・・高精度（ファイン）ピッチサーチ部

107・・・有声音／無声音（V／UV）判別部

108・・・振幅再評価部

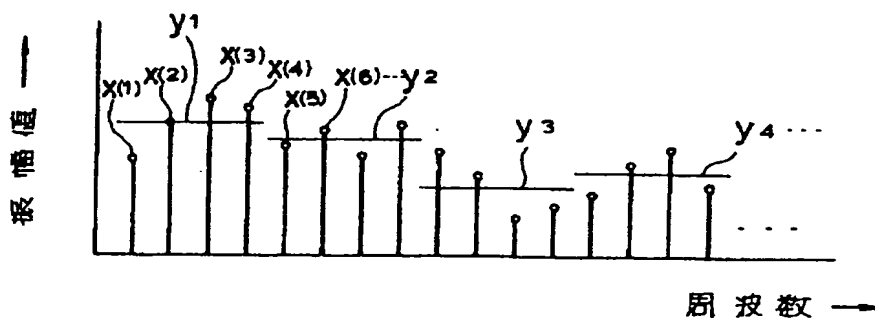
109・・・データ数変換（データレートコンバータ）部

110・・・ベクトル量子化部

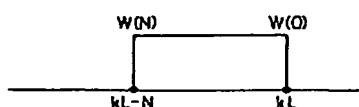
126・・・有声音合成部

127・・・無声音合成部

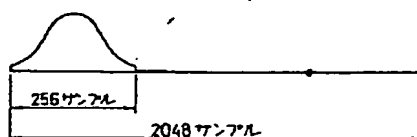
【図4】



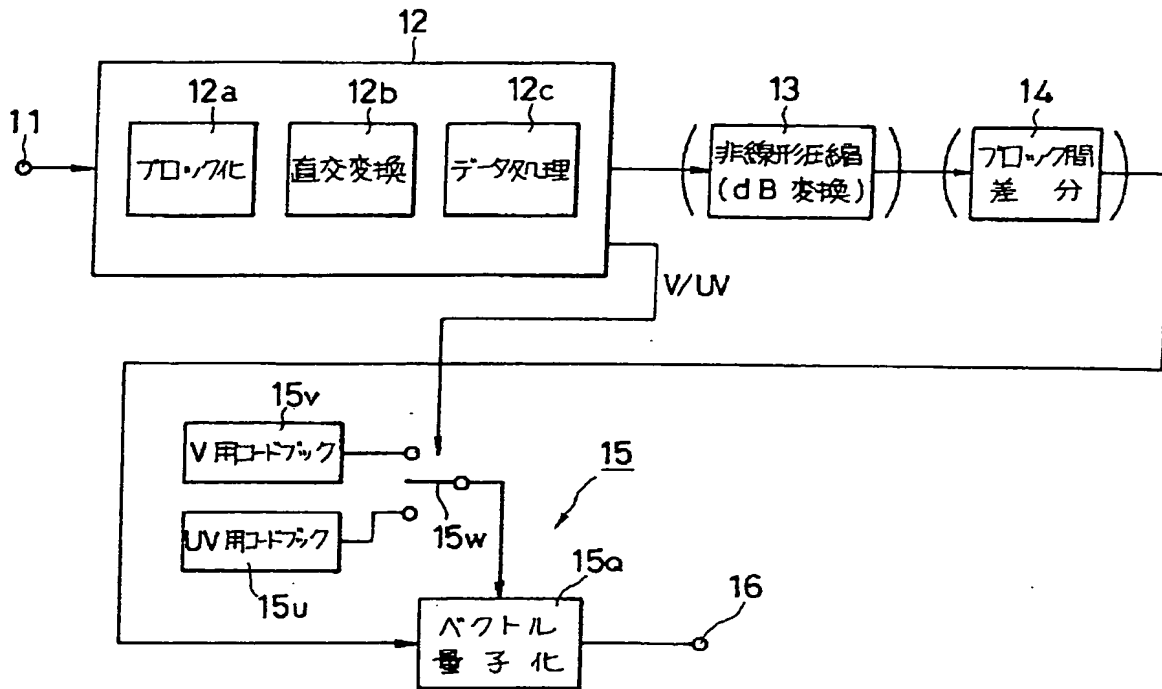
【図12】



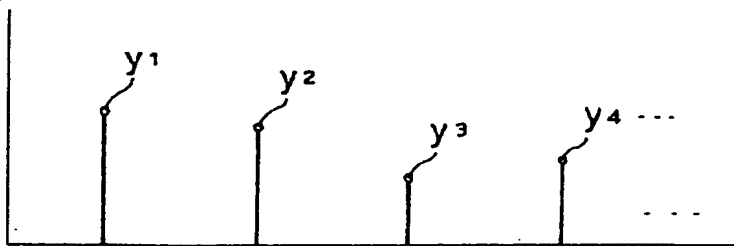
【図13】



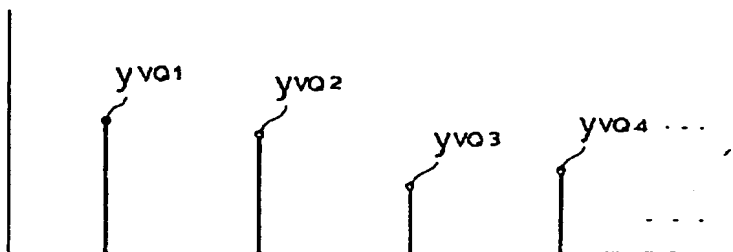
【図1】



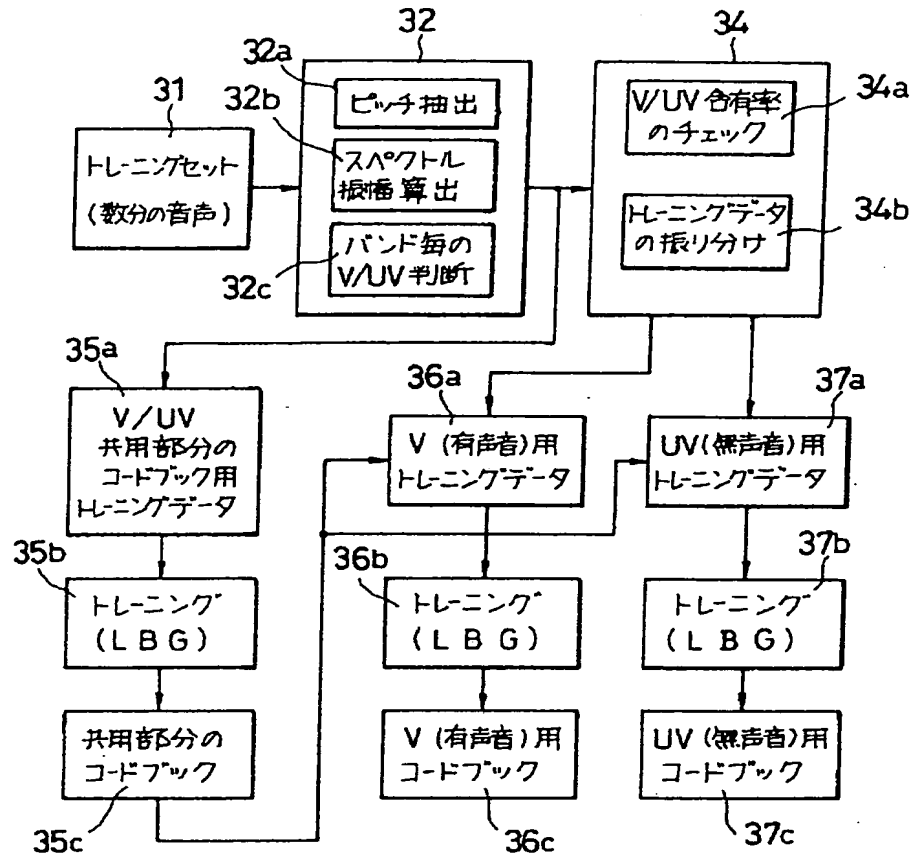
【図5】



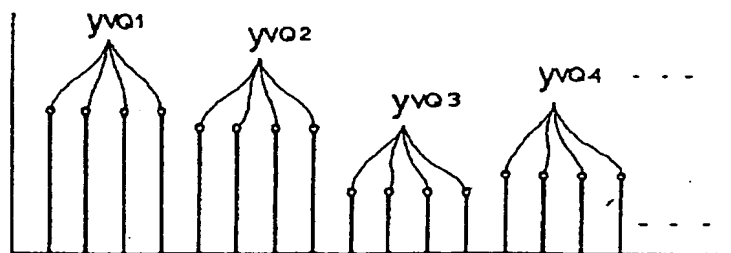
【図6】



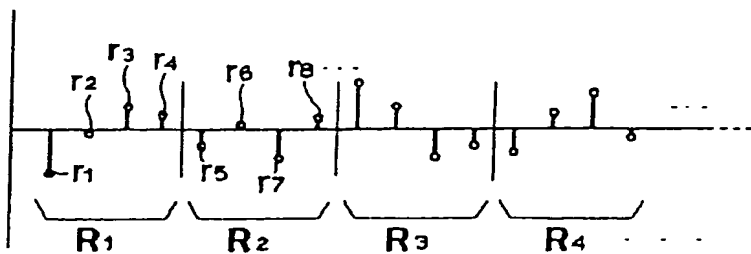
【図2】



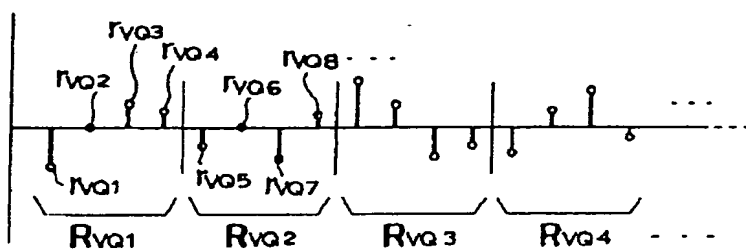
【図7】



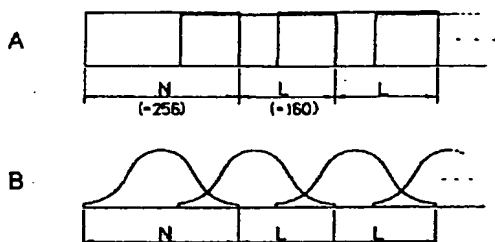
【図8】



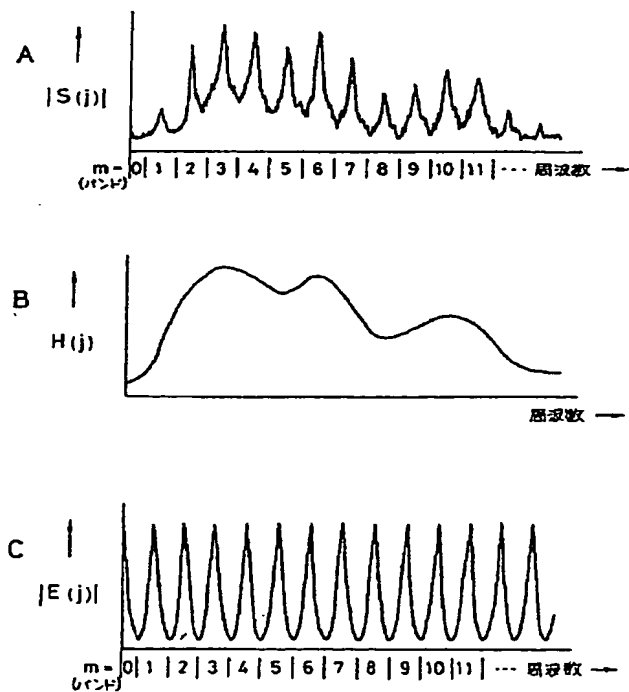
【図9】



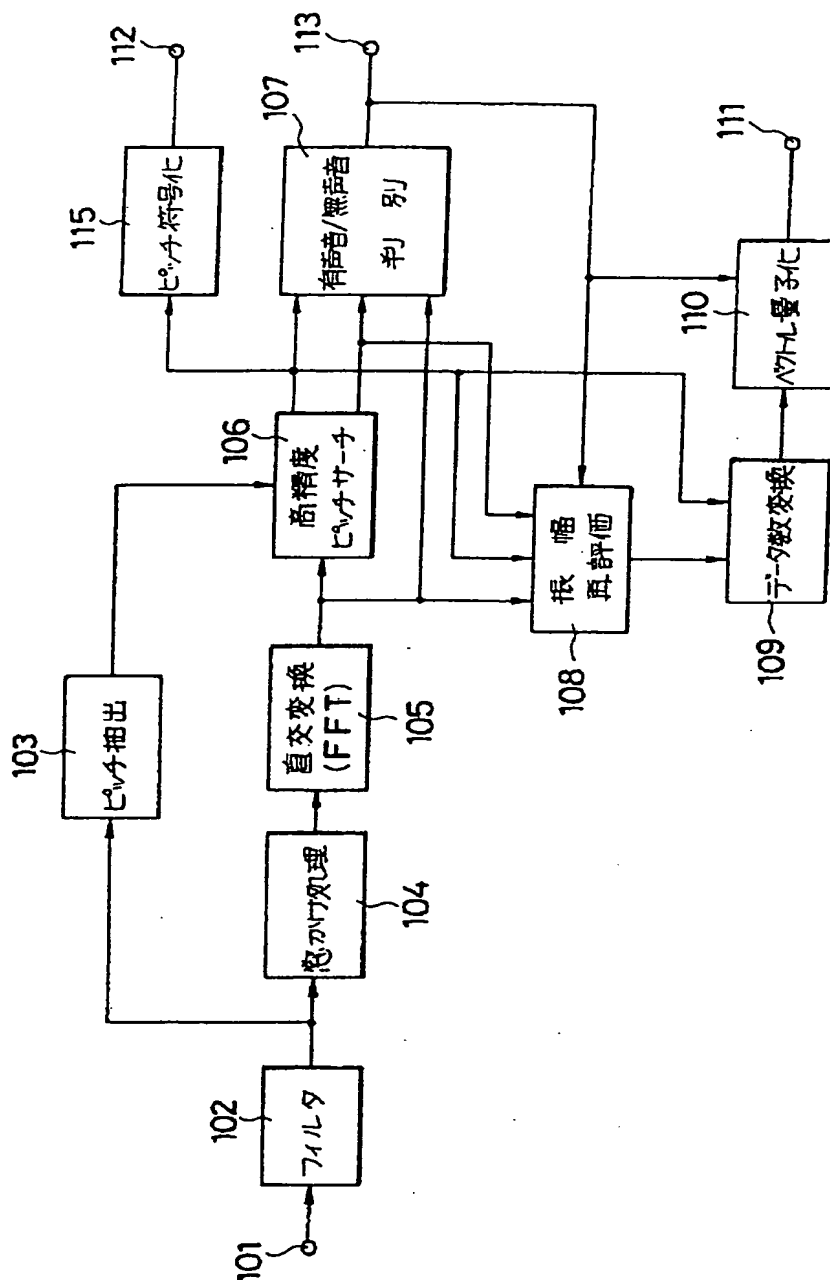
【図11】



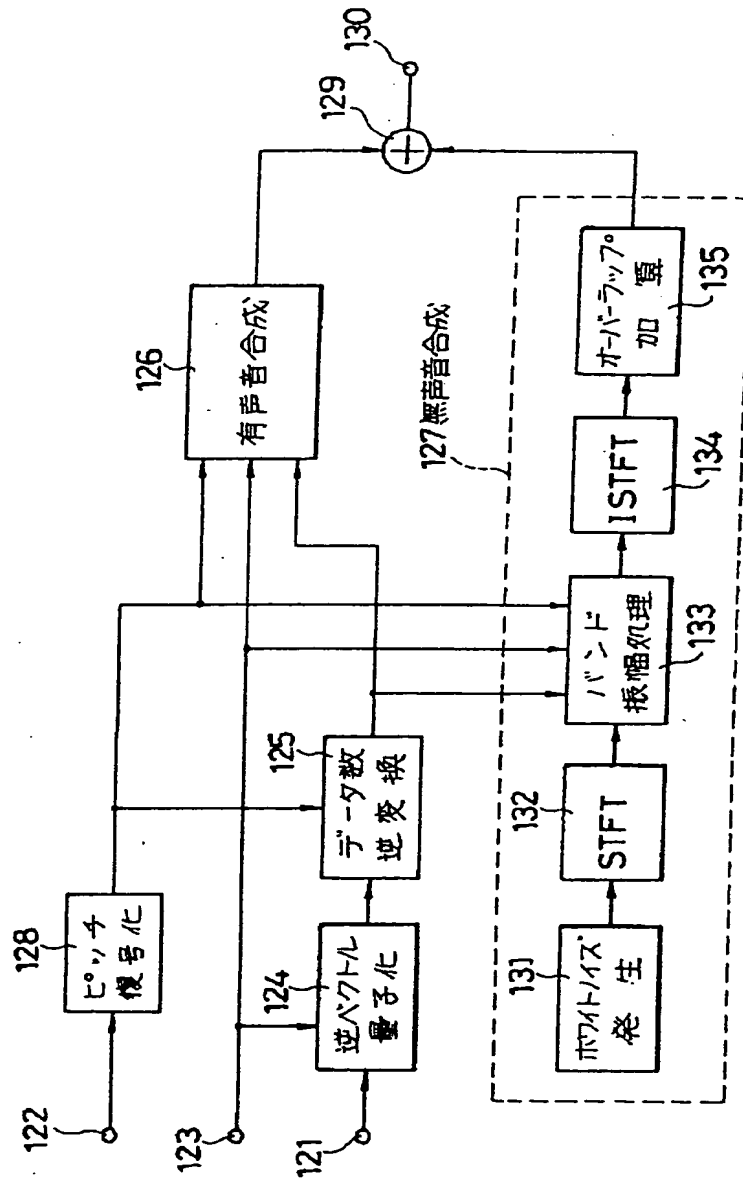
【図14】



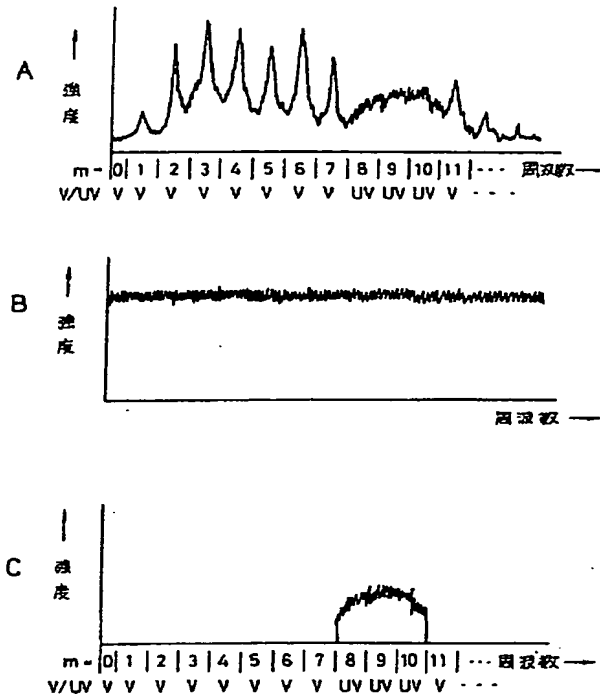
【図10】



〔図15〕



【図16】



【手続補正書】

【提出日】平成5年6月8日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正内容】

【書類名】明細書

【発明の名称】高能率符号化方法

【特許請求の範囲】

【請求項1】 入力されたオーディオ信号をブロック単位で区分して周波数軸に変換して得られるデータに基づいてM次元ベクトルとしての周波数軸上データを求める工程と、

このM次元ベクトルの周波数軸上データに対してベクトル量子化を施すためにオーディオ信号の状態に応じた複数のコードブックを有するベクトル量子化器を用い、上記オーディオ入力信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切り換えて量子化を施す工程とを有することを特徴とする高能率符号化方法。

【請求項2】 上記オーディオ信号として音声信号を用い、上記コードブックとして音声信号が有声音か無声音

かに応じた複数のコードブックを用い、上記特徴パラメータとして上記ブロック毎の入力音声信号が有声音か無声音かを表すパラメータを用いることを特徴とする請求項1記載の高能率符号化方法。

【請求項3】 入力されたオーディオ信号をブロック単位で区分して周波数軸に変換して得られるデータに基づいてM次元ベクトルとしての周波数軸上データを求める工程と、

このM次元ベクトルの周波数軸上データを複数グループに分割して各グループ毎に代表値を求めることにより次元をS次元 ($S < M$) に低下させる工程と、

このS次元ベクトルのデータに対して第1のベクトル量子化を施す工程と、

この第1のベクトル量子化出力データを逆量子化して対応するS次元のコードベクトルを求める工程と、

このS次元のコードベクトルを元のM次元のベクトルに拡張する工程と、

この拡張されたM次元のベクトルと元の上記M次元ベクトルの周波数軸上データとの関係を表すデータに対して第2のベクトル量子化を施すためにオーディオ信号の状態に応じた複数のコードブックを有する第2のベクトル量子化器を用い、上記オーディオ入力信号のブロック毎

の特徴を表すパラメータに応じて上記複数のコードブックを切り換えて量子化を施す工程とを有することを特徴とする高能率符号化方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、入力された音声信号や音響信号等のオーディオ信号をフレーム単位で区分して周波数軸上のデータに変換して符号化を施すような高能率符号化方法に関する。

【0002】

【従来の技術】オーディオ信号（音声信号や音響信号を含む）の時間領域や周波数領域における統計的性質と人間の聴感上の特性を利用して信号圧縮を行うような符号化方法が種々知られている。この符号化方法としては、大別して時間領域での符号化、周波数領域での符号化、分析合成符号化等が挙げられる。

【0003】音声信号等の高能率符号化の例として、MBE（Multiband Excitation: マルチバンド励起）符号化、SBE（Singleband Excitation: シングルバンド励起）符号化、ハーモニック（Harmonic）符号化、SBC（Sub-band Coding: 帯域分割符号化）、LPC（Linear Predictive Coding: 線形予測符号化）、あるいはDCT（離散コサイン変換）、MDCT（モデファイDCT）、FFT（高速フーリエ変換）等において、スペクトル振幅やそのパラメータ（LSPパラメータ、 α パラメータ、 k パラメータ等）のような各種情報データを量子化する場合に、従来においてはスカラ量子化を行うことが多い。

【0004】

【発明が解決しようとする課題】ところで、ビットレートを例えば3~4 kbps程度にまで低減し、量子化効率を更に向上させようとすると、スカラ量子化では量子化雑音（歪み）が大きくなってしまい、実用化が困難であった。そこで、これらの符号化の際に得られる時間軸データや周波数軸データやフィルタ係数データ等を個々に量子化せず、複数のデータを組（ベクトル）にまとめて一つの符号で表現して量子化するベクトル量子化が注目されている。

【0005】ここで、上記MBE、SBE、LPC等のスペクトルエンベロープ、あるいはそのパラメータ（LSPパラメータ、 α パラメータ、 k パラメータ等）のベクトル量子化の際には、固定のコードブックを使用している。しかしながら、使用可能なビット数が減少（低ビットレート化）してくると、固定コードブックでは十分な性能が得られなくなる。このため、ベクトル量子化する入力データのベクトル空間上での存在領域が狭くなるようにクラスタリング（分類分け）したものをベクトル量子化することが好ましい。

【0006】また、伝送ビットレートに余裕があるときも、コードブックサイズ、サーチのための演算量を低減

するため、構造化されたコードブックを使用することも考えられているが、このとき、出力インデックス長が $n+1$ ビットの1個のコードブックを使用する代わりに、例えば2個の n ビットのコードブックに分割する方が好ましい。

【0007】本発明は、このような実情に鑑みてなされたものであり、ベクトル量子化を入力データの性質等に応じて効率良く行え、ベクトル量子化器のコードブックのサイズやサーチ時の演算量を低減でき、品質の高い符号化が行えるような高能率符号化方法の提供を目的とする。

【0008】

【課題を解決するための手段】本発明に係る高能率符号化方法は、入力されたオーディオ信号（音声信号や音響信号等）をブロック単位で区分して周波数軸に変換して得られるデータに基づいて M 次元ベクトルとしての周波数軸上データを求める工程と、この M 次元ベクトルの周波数軸上データに対してベクトル量子化を施すためにオーディオ信号の状態に応じた複数のコードブックを有するベクトル量子化器を用い、上記オーディオ入力信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切り換えて量子化を施す工程とを有することにより、上述の課題を解決するものである。

【0009】他の発明に係る高能率符号化方法としては、入力されたオーディオ信号をブロック単位で区分して周波数軸に変換して得られるデータに基づいて M 次元ベクトルとしての周波数軸上データを求める工程と、この M 次元ベクトルの周波数軸上データを複数グループに分割して各グループ毎に代表値を求めることにより次元を S 次元（ $S < M$ ）に低下させる工程と、この S 次元ベクトルのデータに対して第1のベクトル量子化を施す工程と、この第1のベクトル量子化出力データを逆量子化して対応する S 次元のコードベクトルを求める工程と、この S 次元のコードベクトルを元の M 次元のベクトルに拡張する工程と、この拡張された M 次元のベクトルと元の上記 M 次元ベクトルの周波数軸上データとの関係を表すデータに対して第2のベクトル量子化を施すためにオーディオ信号の状態に応じた複数のコードブックを有する第2のベクトル量子化器を用い、上記オーディオ入力信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切り換えて量子化を施す工程とを有することにより、上述の課題を解決する。

【0010】これらの発明において、上記オーディオ信号として音声信号を用いる場合に、上記コードブックとして音声信号が有声音か無声音かに応じた複数のコードブックを用い、上記特徴パラメータとして上記ブロック毎の入力音声信号が有声音か無声音かを表すパラメータを用いることができる。また、特徴パラメータとして、ピッチの値、ピッチの成分の強弱、有声音/無声音の含有比率、信号スペクトルの傾き及びそのレベル等が使用

でき、基本的には有声音か無声音かに応じてコードブックを切り換えることが好ましい。このような特徴パラメータは、別途伝送してもよいが、符号化方式により予め規定されているような元々伝送されるパラメータを代用させるようにしてもよい。また、上記M次元ベクトルの周波数軸上データとして、上記ブロック単位で周波数軸に変換したデータを非線形圧縮したものを用いることができる。さらに、上記ベクトル量子化の前に、ベクトル量子化しようとするデータのブロック間の差分をとり、このブロック間差分データに対してベクトル量子化を施すようにしてもよい。

【0011】

【作用】入力されたオーディオ信号のブロック毎の特徴を表すパラメータに応じて上記複数のコードブックを切り換えてベクトル量子化を施すことにより、量子化が効率良く行え、ベクトル量子化器のコードブックのサイズやサーチ時の演算量を低減でき、品質の高い符号化が行える。

【0012】

【実施例】以下、本発明に係る高能率符号化方法の実施例について、図面を参照しながら説明する。図1は本発明の一実施例となる高能率符号化方法を説明するための符号化装置（エンコーダ）の概略構成を示している。

【0013】図1において、入力端子11には、音声信号あるいは音響信号が供給され、この入力信号は、周波数軸変換処理部12にて周波数軸上のスペクトル振幅データに変換される。この周波数軸変換処理部12の内部には、例えば入力された時間軸上の信号の所定数のサンプル（Nサンプル）毎にブロック化するブロック化部12a、FFT（高速フーリエ変換）等の直交変換部12b、スペクトルエンベロープの特徴を表す振幅情報を求めるためのデータ処理部12c等が設けられている。周波数軸変換処理部12からの出力は、必要に応じて例えばdB領域に変換する非線形圧縮部13を介し、必要に応じてブロック間の差分をとる処理部14を介してベクトル量子化部15に送られる。このベクトル量子化部15

$$\frac{N_v}{N_v + N_{uv}} \geq V_{th} \quad \dots (1)$$

のときは、V（有声音）用のコードブック15_vを、切り換選択するようにし、それ以外のときはUV（無声音）用コードブック15_{uv}を切り換選択するようにする。上記閾値V_{th}は、例えば1程度に設定すればよい。

【0017】デコーダ（合成系）側でも同様に、V（有声音）用及びUV（無声音）用の2種類のコードブックの切り換選択を行う。上記V/UV判別フラグはMBEポコーダでは必ず伝送するサイドインフォメーション（補助情報）であるので、この具体例においては別個にコードブック切り換のための特徴パラメータを伝送する必要が

なく、伝送ビットレートの増加がない。
*5では、入力データの所定数（Mサンプル）がまとめられてM次元とベクトルとされ、ベクトル量子化処理が施される。このようなM次元ベクトル量子化処理は、一般的には、入力されたM次元ベクトルに対してM次元空間上で最も距離に近いコードベクトルがコードブックから検索（サーチ）されて、該検索されたコードベクトルのインデックスを出力端子16から取り出すような処理であるが、この図1に示す実施例のベクトル量子化部15は、複数種類のコードブックを有し、これらのコードブックが、周波数軸変換処理部12からの上記入力信号の特徴を表すパラメータに応じて切り換えられるようになっている。

【0014】図1の例では、入力信号として音声信号を想定し、V（有声音）用コードブック15_vと、UV（無声音）用コードブック15_{uv}とが切換スイッチ15により切り換えられてベクトル量子化器15に送られるようになっており、切換スイッチ15は、周波数軸変換処理部12からのV/UV（有声音/無声音）判別信号に応じて切換制御される。このV/UV判別信号（フラグ等）は、後述するようなMBE（Multiband Excitation: マルチバンド励起）ポコーダ（音声分析合成装置）等の場合には、分析系（エンコーダ）側から合成系（デコーダ）側へ伝送されるパラメータであり、別途に伝送する必要がない。

【0015】ここで上記MBEの場合を例にとって説明すると、上記コードブック15_v、15_{uv}の切り換のためのパラメータは、伝送される情報の一つであるV/UV判別フラグを利用するようにすればよい。すなわち、上記周波数軸変換処理部12において、ピッチに応じたバンド分割を行い、分割された各バンド毎にV（有声音）かUV（無声音）かの判別が行われる。ここで、Vバンドの個数をN_v、UVバンドの個数をN_{uv}とし、所定の閾値V_{th}に対して、

【0016】

【数1】

なく、伝送ビットレートの増加がない。

【0018】V（有声音）用コードブック15_v、UV（無声音）用コードブック15_{uv}の生成（トレーニング）は、単に同じ基準でトレーニングデータを分けることによって可能となる。すなわち、V（有声音）と判別された振幅データの集まりから生成されたコードブックをV用コードブック15_vとし、UV（無声音）と判別された振幅データの集まりから生成されたコードブックをUV用コードブック15_{uv}とする。

【0019】この具体例においては、V/UV情報をコ

ードブックの切換に用いていることから、上記V/UV判別フラグをより確実に(信頼性を高く)しておく必要がある。例えば、明らかに子音や背景雑音であるとみなせる部分では、全バンドUVにすべきである。その判断の一例としては、高域にパワーの多い微小入力をUV化することが挙げられる。

【0020】入力信号のN点(256サンプル)に対し*

$$\left. \begin{aligned} P_L &= \sum_{i=0}^{(N/4)-1} \text{rms}^2(i) \\ P_H &= \sum_{i=N/4}^{(N/2)-1} \text{rms}^2(i) \end{aligned} \right\} \dots (2)$$

ただし、 $\text{rms}(i)$ は、 $\sqrt{\text{Re}^2(i) + \text{Im}^2(i)}$
 $\text{Re}(i)$ 及び $\text{Im}(i)$ は入力数列のFFTの実部
 及び虚部である。

この(2)式の P_L 、 P_H を用いて、
 【0022】

※【数3】

$$\left. \begin{aligned} R_d &= \frac{P_L}{P_H} \\ L &= \sqrt{\frac{P_L + P_H}{N/2}} \end{aligned} \right\} \dots (3)$$

として、 $R_d < R_{th}$ 、かつ $L < L_{th}$ のときは、無条件に全バンドUV(無声音)と判別する。

【0023】これを行うと、間違ったピッチを微小入力で検出したときに、それを使用しない効果がある。このようにして、より確かなV/UVフラグを作っておくと、ベクトル量子化の際のコードブックの切換に好都合である。

【0024】次に、上記V(有声音)用及びUV(無声音)用の各コードブックを作成する際のトレーニングについて、図2を参照しながら説明する。図2において、トレーニング用の数分程度の音声信号から成るトレーニングセット31からの信号は、周波数軸変換処理部32

*てFFT(高速フーリエ変換)を行い、有効な $0 \sim \pi$ ($0 \sim N/2$)の中で、 $0 \sim N/4$ と $N/4 \sim N/2$ とのそれぞれの区間でパワー計算を行う。

【0021】

【数2】

に送られて、ピッチ抽出部32aによりピッチ抽出が、スペクトル振幅算出部32bによりスペクトル振幅の算出が、またバンド毎のV/UV判断部32cにより各バンド毎のV/UV判断がそれぞれ行われる。この周波数軸変換処理部32からの出力データは、トレーニング前処理部34に送られる。

【0025】トレーニング前処理部34では、V/UV含有率のチェック部34aにより上記(1)式や(4)式の条件等をチェックし、得られたV/UV情報に応じて、トレーニングデータの振り分け部34bにより上記スペクトル振幅データを振り分けており、V(有声音)のときにはV(有声音)用トレーニングデータ出力部3

6aへ、UV（無声音）のときにはUV（有声音）用トレーニングデータ出力部37aへ、それぞれ振幅データを送っている。

【0026】V用トレーニングデータ出力部36aから出力されたV（有声音）のスペクトル振幅データは、トレーニング処理部36bに送られて、例えばいわゆるLBG法によりトレーニング処理が行われ、V（有声音）用コードブック36cが作成される。ここで、上記LBG法とは、Linde, Y., Buzo, A. and Gray, R. M., "An Algorithm for Vector Quantizer Design", IEEE Trans. Comm., COM-28, pp.84-95, Jan. 1980 において提案されたベクトル量子化器を設計するアルゴリズムにおけるコードブックのトレーニング法であり、確率密度関数が知られていない情報源に対していわゆるトレーニング系列を用いて最適なベクトル量子化器を設計するものである。同様に、UV用トレーニングデータ出力部37aから出力されたUV（無声音）のスペクトル振幅データは、トレーニング処理部37bに送られて例えばLBG法によりトレーニング処理が行われ、UV（無声音）用コードブック37cが作成される。

【0027】ここで、後述するように、ベクトル量子化部に階層構造を持たせ、上位層にはV/UV共用部分のコードブックを用い、下位層のコードブックのみをV/UVに応じて切り換えるようにする場合には、V/UV共用部分のコードブックも作成する必要がある。この場合、上記周波数軸変換処理部32からの出力データをV/UV共用部分のコードブック用トレーニングデータ出力部35aに送ることが必要とされる。

【0028】V/UV共用部分のコードブック用トレーニングデータ出力部35aから出力されたスペクトル振幅データは、トレーニング処理部35bに送られて、例えばいわゆるLBG法によりトレーニング処理が実行され、V/UV共用部分のコードブック35cが作成される。この作成されたV/UV共用コードブック35cからのコードベクトルを、上記V用、UV用の各トレーニングデータ出力部36a、37aに送り、V/UV共用コードブックを用いて、V用及びUV用のトレーニングデータに対し上位層のベクトル量子化を実行し、下位層のためのV用及びUV用のトレーニングデータを生成する必要がある。

【0029】以下、階層構造化されたベクトル量子化部の具体的構成及び作用について、図3～図9を参照しながら説明する。すなわち、この図3に示すベクトル量子化部15は、例えば上下の2つの層に階層構造化されており、入力ベクトルに対して2段階のベクトル量子化が施されるようになっている。

【0030】図3に示すベクトル量子化器15の入力端子17には、上記図1の周波数軸変換処理部12からの周波数軸上の振幅データが（必要に応じて非線形圧縮部13及びブロック間差分処理部14を介し）、ベクトル

量子化の単位となる上記M次元ベクトルとして供給されている。このM次元ベクトルは、次元低減部21に送られることにより複数グループに分割され、各グループ毎に代表値を求めることにより次元がS次元（ $S < M$ ）に低下させられる。ここで図4は、ベクトル量子化器15に入力されるM次元ベクトルの各要素、すなわち周波数軸上のM個の振幅データ $x(n)$ の一具体例を示しており、 $1 \leq n \leq M$ である。これらのM個の振幅データ $x(n)$ は、例えば4サンプル毎にまとめられてそれぞれの代表値、例えば平均値 y 、が求められ、図5に示すように、平均値データ $y_1 \sim y_S$ のS個（この場合は $S = M/4$ ）から成るS次元ベクトルが得られる。

【0031】次に、このS次元ベクトルのデータに対してS次元ベクトル量子化器22によりベクトル量子化が施される。すなわち、S次元ベクトル量子化器22のコードブック22。中のS次元コードベクトルの内、入力されたS次元ベクトルにS次元空間上で最も距離が近いコードベクトルがサーチされ、このサーチされたコードベクトルのインデックスデータは出力端子26より取り出され、サーチされたコードベクトル（出力インデックスを逆ベクトル量子化して得られるコードベクトル）は次元拡張部23に送られる。コードブック22。には、上記図2において説明したV/UV共用部分のコードブック35cが用いられる。図6は、上記図5に示すS個の平均値データ $y_1 \sim y_S$ から成るS次元ベクトルをベクトル量子化した後、逆量子化して（あるいはベクトル量子化器22のコードブックで上記量子化の際にサーチされたコードベクトルを取り出して）得られたローカルデコーダ出力としてのS次元ベクトル v_q の各要素 $y_{vq1} \sim y_{vqs}$ を示している。

【0032】次に次元拡張部23では、上記S次元のコードベクトルを元のM次元のベクトルに拡張する。この拡張されたM次元ベクトルの各要素の例を図7に示す。この図7から明らかなように、上記逆ベクトル量子化されたS次元ベクトル v_q の各要素 $y_{vq1} \sim y_{vqs}$ をそれぞれ元の4サンプルずつに増加させることにより、 $4S = M$ 個の要素から成るM次元ベクトルを得るわけである。この拡張されたM次元のベクトルと元の上記M次元ベクトルの周波数軸上データとの関係を表すデータに対して第2のベクトル量子化を施す。

【0033】図3の具体例では、次元拡張部23からの拡張されたM次元ベクトルデータを減算器24に送り、元の上記M次元ベクトルの周波数軸上データから減算することにより、上記S次元を拡張したM次元ベクトルと元のM次元ベクトルとの関係を表すS個のベクトルデータを得ている。図8は、上記図4に示すM次元ベクトルの各要素である周波数軸上のM個の振幅データ $x(n)$ から、図7に示す拡張M次元ベクトルの各要素を減算して得られたM個のデータ $r_1 \sim r_M$ を示しており、これらのM個のデータ $r_1 \sim r_M$ の4サンプルずつを組

(ベクトル)としてS個の4次元ベクトル $\mathbf{v}_{q1} \sim \mathbf{v}_{qS}$ が得られる。

【0034】このようにして減算器24から得られたS個のベクトルのそれぞれに対して、ベクトル量子化器群25のS個の各ベクトル量子化器25_{1q}～25_{Sq}によりそれぞれベクトル量子化を施す。各ベクトル量子化器25_{1q}～25_{Sq}から出力されるインデックスは、それぞれ出力端子27_{1q}～27_{Sq}から取り出される。図9は、ベクトル量子化器25_{1q}～25_{Sq}としてそれぞれ4次元ベクトル量子化器を用い、上記図8に示す各4次元ベクトル $\mathbf{v}_{q1} \sim \mathbf{v}_{qS}$ をそれぞれベクトル量子化した後の各4次元ベクトル $\mathbf{v}_{q1} \sim \mathbf{v}_{qS}$ の各要素 $r_{vq1} \sim r_{vqS}$ 、 $r_{vq1} \sim r_{vqS}$ 、 $\dots \sim r_{vqS}$ を示している。

【0035】これらの各ベクトル量子化器25_{1q}～25_{Sq}には、それぞれV(有声音)用のコードブック25_{1v}～25_{Sv}とUV(無声音)用のコードブック25_{1u}～25_{Su}とが用いられ、これらのV用コードブック25_{1v}～25_{Sv}とUV用コードブック25_{1u}～25_{Su}とは、入力端子18からのV/UV情報に応じて切換制御される切換スイッチ25_{1v}～25_{Sv}により切換選択されるようになっている。これらの切換スイッチ25_{1v}～25_{Sv}の切換制御は、全バンドに対して同時に(連動させて)行ってもよいが、各ベクトル量子化器25_{1q}～25_{Sq}が受け持つ周波数帯域が異なることを考慮して、バンド毎のV/UV判別フラグに応じて切換制御するようにしてもよい。V用コードブック25_{1v}～25_{Sv}が図2のV(有声音)用コードブック36cに対応し、UV用コードブック25_{1u}～25_{Su}が図2のV(有声音)用コードブック37cに対応することは勿論である。

【0036】このような階層構造化された2段階のベクトル量子化を施すことにより、コードブックサーチのための演算量を低減でき、コードブックのためのメモリ量(例えばROM容量)を低減でき、また上記出力端子26から得られる上位層のより重要なインデックスに対して誤り訂正符号化を施して重点的に保護すること等により、誤り訂正符号の効果的な適用が可能となる。なお、ベクトル量子化部15の階層構造は、2段階に限定されず、3段階以上の多層の階層構造を持たせるようにしてもよい。

【0037】なお、図1～図3の各部は、全てをハードウェアにて構成する必要はなく、例えばいわゆるDSP(デジタル信号プロセッサ)等を用いてソフトウェア的に実現するようにしてもよい。

【0038】以上説明したように、例えば音声合成分析符号化の場合に、音声の有声無声の度合い、ピッチ等が既に特徴量として抽出されていることを考慮して、それらの特徴量、特に有聲/無声の判断結果によってベクトル量子化のコードブックを切り換えることにより、良好なベクトル量子化が実現できる。すなわち、有声音時と無声音時とでは、各々存在するスペクトルの形状が大き

く異なり、各状態に対応して各々別々にトレーニングされたコードブックを持つことは、特性改善上、大変好ましいことである。また、階層構造化ベクトル量子化の場合、上位層のベクトル量子化は固定のコードブックとし、下位層のベクトル量子化のコードブックのみ、有声音/無声音の2種を切り換えるようにしてもよい。また、周波数軸上のビットアロケーションも、例えば有声音のときは低音重視、無声音のときは高音重視に切り換えるようにしてもよい。切換制御には、ピッチの有無、有声音/無声音の比率、スペクトルのレベルや傾き、等を利用することができる。さらに、3種類以上のコードブックを切り換えるようにしてもよく、例えば、無声音用のコードブックを子音、背景雑音等で2種類以上用いるようにしてもよい。

【0039】次に、上述したような高能率符号化方法が適用可能な、音声信号の合成分析符号化装置(いわゆるボコーダ)の一種のMBE(Multiband Excitation: マルチバンド励起)ボコーダの具体例について、図面を参照しながら説明する。以下に説明するMBEボコーダは、D.W. Griffin and J.S. Lim, "Multiband Excitation Vocoder," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36, No. 8, pp. 1223-1235, Aug. 1988に開示されているものであり、従来のPARCOR(Partial auto-CORrelation: 偏自己相関)ボコーダ等では、音声のモデル化の際に有声音区間と無声音区間とをブロックあるいはフレーム毎に切り換えていたのに対し、MBEボコーダでは、同時刻(同じブロックあるいはフレーム内)の周波数軸領域に有声音(Voiced)区間と無声音(Unvoiced)区間とが存在するという仮定でモデル化している。

【0040】図10は、上記MBEボコーダに本発明を適用した実施例の全体の概略構成を示すブロック図である。この図10において、入力端子101には音声信号が供給されるようになっており、この入力音声信号は、HPF(ハイパスフィルタ)等のフィルタ102に送られて、いわゆるDC(直流)オフセット分の除去や帯域制限(例えば200～3400Hzに制限)のための少なくとも低域成分(200Hz以下)の除去が行われる。このフィルタ102を介して得られた信号は、ピッチ抽出部103及び窓かけ処理部104にそれぞれ送られる。ピッチ抽出部103では、入力音声信号データが所定サンプル数N(例えばN=256)単位でブロック分割され(あるいは方形窓による切り出しが行われ)、このブロック内の音声信号についてのピッチ抽出が行われる。このような切り出しブロック(256サンプル)を、例えば図11のAに示すようにLサンプル(例えばL=160)のフレーム間隔で時間軸方向に移動させており、各ブロック間のオーバーラップはN-Lサンプル(例えば96サンプル)となっている。また、窓かけ処理部104では、1ブロックNサンプルに対して所定の窓関数、

例えばハミング窓をかけ、この窓かけブロックを1フレームLサンプルの間隔で時間軸方向に順次移動させてい*

$$x_s(k, q) = x(q) w(kL - q)$$

となる。この(4)式において、kはブロック番号を、qはデータの時間インデックス(サンプル番号)を表し、処理前の入力信号のq番目のデータx(q)に対して第kブロックの窓(ウィンドウ)関数w(kL-q)により窓*

$$w_s(r) = \begin{cases} 1 & 0 \leq r < N \\ 0 & r < 0, N \leq r \end{cases} \quad \dots (5)$$

また、窓かけ処理部104での図11のBに示すような★ ★ハミング窓の場合の窓関数w_s(r)は、

$$w_s(r) = \begin{cases} 0.54 - 0.46 \cos(2\pi r/(N-1)) & 0 \leq r < N \\ 0 & r < 0, N \leq r \end{cases} \quad \dots (6)$$

である。このような窓関数w_s(r)あるいはw_s(r)を用いるときの上記(4)式の窓関数w(r) (=w(kL-q))の非零区間は、

$$0 \leq kL - q < N$$

これを変形して、

$$kL - N < q \leq kL$$

従って例えば上記方形窓の場合に窓関数w_s(kL-q) = 1となるのは、図12に示すように、kL-N < q ≤ kLのときとなる。また、上記(4)～(6)式は、長さN (=256) サンプルの窓が、L (=160) サンプルずつ前進してゆくことを示している。以下、上記(5)式、(6)式の各窓関数で切り出された各N点(0 ≤ r < N)の非零サンプル列を、それぞれx_s(k, r)、x_s(k, r)と表すことにする。

【0042】窓かけ処理部104では、図13に示すように、上記(6)式のハミング窓がかけられた1ブロック256サンプルのサンプル列x_s(k, r)に対して1792サンプル分の0データが付加されて(いわゆる0詰めされて)2048サンプルとされ、この2048サンプルの時間軸データ列に対して、直交変換部105により例えばFFT(高速フーリエ変換)等の直交変換処理が施される。

【0043】ピッチ抽出部103では、上記x_s(k, r)のサンプル列(1ブロックNサンプル)に基づいてピッチ抽出が行われる。このピッチ抽出法には、時間波形の周期性や、スペクトルの周期的周波数構造や、自己相関関数を用いるもの等が知られているが、本実施例では、センタクリップ波形の自己相関法を採用している。このときのブロック内でのセンタクリップレベルについては、1ブロックにつき1つのクリップレベルを設定してもよいが、ブロックを細分割した各部(各サブブロック)の信号のピークレベル等を検出し、これらの各サブブロックのピークレベル等の差が大きいために、ブロック内でクリップレベルを段階的にあるいは連続的に変化☆

$$S(j) = H(j) \mid E(j) \mid$$

と表現するようなモデルを想定している。ここで、Jは $\omega_s / 4\pi = f_s / 2$ に対応し、サンプリング周波数f

＊る。

【0041】このような窓かけ処理を数式で表すと、

$$\dots (4)$$

＊かけ処理されることによりデータx_s(k, q)が得られることを示している。ピッチ抽出部103内での図11のAに示すような方形窓の場合の窓関数w_s(r)は、

☆させるようにしている。このセンタクリップ波形の自己相関データのピーク位置に基づいてピッチ周期を決めている。このとき、現在フレームに属する自己相関データ(自己相関は1ブロックNサンプルのデータを対象として求められる)から複数のピークを求めておき、これらの複数のピークの内の最大ピークが所定の閾値以上のときには該最大ピーク位置をピッチ周期とし、それ以外のときには、現在フレーム以外のフレーム、例えば前後のフレームで求められたピッチに対して所定の関係を満たすピッチ範囲内、例えば前フレームのピッチを中心として±20%の範囲内にあるピークを求め、このピーク位置に基づいて現在フレームのピッチを決定するようにしている。このピッチ抽出部103ではオープンループによる比較的ラフなピッチのサーチが行われ、抽出されたピッチデータは高精度(ファイン)ピッチサーチ部106に送られて、クローズドループによる高精度のピッチサーチ(ピッチのファインサーチ)が行われる。

【0044】高精度(ファイン)ピッチサーチ部106には、ピッチ抽出部103で抽出された整数(インテジャ)値の粗(ラフ)ピッチデータと、直交変換部105により例えばFFTされた周波数軸上のデータとが供給されている。この高精度ピッチサーチ部106では、上記粗ピッチデータ値を中心に、0.2～0.5きざみで±数サンプルずつ振って、最適な小数点付き(フローティング)のファインピッチデータの値へ追い込む。このときのファインサーチの手法として、いわゆる合成による分析(Analysis by Synthesis)法を用い、合成されたパワースペクトルが原音のパワースペクトルに最も近くなるようにピッチを選んでいる。

【0045】このピッチのファインサーチについて説明する。先ず、上記MBEボコーダにおいては、上記FFT等により直交変換された周波数軸上のスペクトルデータとしてのS(j)を

$$0 < j < J \quad \dots (7)$$

、 $\omega_s / 2\pi$ が例えば8 kHzのときには4 kHzに対応する。上記(7)式中において、周波数軸上のスペクト

ルデータ $S(j)$ が図14のAに示すような波形のとき、 $H(j)$ は、図14のBに示すような元のスペクトルデータ $S(j)$ のスペクトル包絡線 (エンベロープ) を示し、 $E(j)$ は、図14のCに示すような等レベルで周期的な励起信号 (エキサイテーション) のスペクトルを示している。すなわち、FFTスペクトル $S(j)$ は、スペクトルエンベロープ $H(j)$ と励起信号のパワースペクトル $E(j)$ との積としてモデル化される。

【0046】上記励起信号のパワースペクトル $E(j)$ は、上記ピッチに応じて決定される周波数軸上の波形の周期性 (ピッチ構造) を考慮して、1つの帯域 (バンド) の波形に相当するスペクトル波形を周波数軸上の各バンド毎に繰り返すように配列することにより形成される。この1バンド分の波形は、例えば上記図13に示す*

* ような256サンプルのハミング窓関数に1792サンプル分の0データを付加 (0詰め) した波形を時間軸信号と見なしてFFTし、得られた周波数軸上のある帯域幅を持つインパルス波形を上記ピッチに応じて切り出すことにより形成することができる。

【0047】次に、上記ピッチに応じて分割された各バンド毎に、上記 $H(j)$ を代表させるような (各バンド毎のエラーを最小化するような) 値 (一種の振幅) $|A_m|$ を求める。ここで、例えば第 m バンド (第 m 高調波の帯域) の下限、上限の点をそれぞれ a_m 、 b_m とするとき、この第 m バンドのエラー ε_m は、

【0048】

【数4】

$$\varepsilon_m = \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \}^2 \dots (18)$$

【0049】で表せる。このエラー ε_m を最小化するよ
うな $|A_m|$ は、

【0050】

※

$$\frac{\partial \varepsilon_m}{\partial |A_m|} = -2 \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \} |E(j)| = 0$$

$$\therefore |A_m| = \frac{\sum_{j=a_m}^{b_m} |S(j)| |E(j)|}{\sum_{j=a_m}^{b_m} |E(j)|^2} \dots (19)$$

【0051】となり、この(6)式の $|A_m|$ のとき、エラー ε_m を最小化する。このような振幅 $|A_m|$ を各バンド毎に求め、得られた各振幅 $|A_m|$ を用いて上記(8)式で定義された各バンド毎のエラー ε_m を求める。次に、このような各バンド毎のエラー ε_m の全バンドの総和値 $\Sigma \varepsilon_m$ を求める。さらに、このような全バンドのエラー総和値 $\Sigma \varepsilon_m$ を、いくつかの微小に異なるピッチについて求め、エラー総和値 $\Sigma \varepsilon_m$ が最小となるようなピッチを求める。

【0052】すなわち、上記ピッチ抽出部103で求められたラフピッチを中心として、例えば0.25きざみで上下に数種類ずつ用意する。これらの複数種類の微小に異なるピッチの各ピッチに対してそれぞれ上記エラー総和値 $\Sigma \varepsilon_m$ を求める。この場合、ピッチが定まるとバンド幅が決まり、上記(9)式より、周波数軸上データのパワースペクトル $|S(j)|$ と励起信号スペクトル $|E(j)|$ とを用いて上記(8)式のエラー ε_m を求め、その全バンドの総和値 $\Sigma \varepsilon_m$ を求めることができる。この

エラー総和値 $\Sigma \varepsilon_m$ を各ピッチ毎に求め、最小となるエラー総和値に対応するピッチを最適のピッチとして決定するわけである。以上のようにして高精度ピッチサーチ部106で最適のファイン (例えば0.25きざみ) ピッチが求められ、この最適ピッチに対応する振幅 $|A_m|$ が決定される。

【0053】以上ピッチのファインサーチの説明においては、説明を簡略化するために、全バンドが有声音 (Voiced) の場合を想定しているが、上述したようにMBEポコーダにおいては、同時刻の周波数軸上に無声音 (Unvoiced) 領域が存在するというモデルを採用していることから、上記各バンド毎に有声音/無声音の判別を行うことが必要とされる。

【0054】上記高精度ピッチサーチ部106からの最適ピッチ及び振幅 $|A_m|$ のデータは、有声音/無声音判別部107に送られ、上記各バンド毎に有声音/無声音の判別が行われる。この判別のために、NSR (ノイズ to シグナル比) を利用する。すなわち、第 m バンド

のNSRは、
【0055】

*【数6】

*

$$NSR = \frac{\sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \}^2}{\sum_{j=a_m}^{b_m} |S(j)|^2} \quad \dots (20)$$

【0056】と表せ、このNSR値が所定の閾値（例えば0.3）より大のとき（エラーが大きい）ときには、そのバンドでの $|A_m| |E(j)|$ による $|S(j)|$ の近似が良くない（上記励起信号 $|E(j)|$ が基底として不適当である）と判断でき、当該バンドをUV（Unvoiced、無声音）と判別する。これ以外のときは、近似がある程度良好に行われていると判断でき、そのバンドをV（Voiced、有声音）と判別する。

【0057】次に、振幅再評価部108には、直交変換部105からの周波数軸上データ、高精度ピッチサーチ※

※部106からのファインピッチと評価された振幅 $|A_m|$ との各データ、及び上記有声音／無声音判別部107からのV／UV（有声音／無声音）判別データが供給されている。この振幅再評価部108では、有声音／無声音判別部107において無声音（UV）と判別されたバンドに関して、再度振幅を求めている。このUVのバンドについての振幅 $|A_m|_{uv}$ は、

【0058】

【数7】

$$|A_m|_{uv} = \sqrt{\sum_{j=a_m}^{b_m} |S(j)|^2 / (b_m - a_m + 1)}$$

... (21)

【0059】にて求められる。

【0060】この振幅再評価部108からのデータは、データ数変換（一種のサンプリングレート変換）部109に送られる。このデータ数変換部109は、上記ピッチに応じて周波数軸上での分割帯域数が異なり、データ数（特に振幅データの数）が異なることを考慮して、一定の個数にするためのものである。すなわち、例えば有効帯域を3400kHzまでとすると、この有効帯域が上記ピッチに応じて、8バンド～63バンドに分割されることになり、これらの各バンド毎に得られる上記振幅 $|A_m|$ （UVバンドの振幅 $|A_m|_{uv}$ も含む）データの個数 $m_{xx}+1$ も8～63と変化することになる。このためデータ数変換部109では、この可変個数 $m_{xx}+1$ の振幅データを一定個数 N_c （例えば44個）のデータに変換している。

【0061】ここで本実施例においては、周波数軸上の有効帯域1ブロック分の振幅データに対して、ブロック内の最後のデータからブロック内の最初のデータまでの値を補間するようなダミーデータを付加してデータ個数を N_c 個に拡大した後、帯域制限型の K_{os} 倍（例えば8倍）のオーバーサンプリングを施すことにより K_{os} 倍の個数の振幅データを求め、この K_{os} 倍の個数（ $(m_{xx}+1) \times K_{os}$ 個）の振幅データを直線補間してさらに多くの N_c 個（例えば2048個）に拡張し、この N_c 個の

データを間引いて上記一定個数 N_c （例えば44個）のデータに変換する。

【0062】このデータ数変換部109からのデータ（上記一定個数 N_c の振幅データ）がベクトル量子化部110に送られて、所定個数のデータ毎にまとめられてベクトルとされ、ベクトル量子化が施される。このベクトル量子化部110として、上述した図1あるいは図3に示すような、V（有声音）用コードブックとUV（無声音）用コードブックとを切換選択するようなベクトル量子化部15を用いている。これらのV用、UV用コードブックは、上記有声音／無声音（V／UV）判別部107からのV／UV判別データに応じて切換制御すればよい。このようなベクトル量子化部110からの量子化出力データは、出力端子111を介して取り出される。

【0063】また、上記高精度のピッチサーチ部106からの高精度（ファイン）ピッチデータは、ピッチ符号化部115で符号化され、出力端子112を介して取り出される。さらに、上記有声音／無声音判別部107からの有声音／無声音（V／UV）判別データは、出力端子113を介して取り出される。これらの各出力端子111～113からのデータは、所定の伝送フォーマットの信号とされて伝送される。

【0064】なお、これらの各データは、上記 N サンプル（例えば256サンプル）のブロック内のデータに対

して処理を施すことにより得られるものであるが、ブロックは時間軸上を上記1サンプルのフレームを単位として前進することから、伝送するデータは上記フレーム単位で得られる。すなわち、上記フレーム周期でピッチデータ、V/U判別データ、振幅データが更新されることになる。

【0065】次に、伝送されて得られた上記各データに基づき音声信号を合成するための合成側（デコード側）の概略構成について、図15を参照しながら説明する。この図15において、入力端子121には上記ベクトル量子化された振幅データが、入力端子122には上記符号化されたピッチデータが、また入力端子123には上記V/U判別データがそれぞれ供給される。入力端子121からの量子化振幅データは、逆ベクトル量子化部124に送られて逆量子化される。この逆ベクトル量子化部124も、V（有声音）用コードブックとUV（無声音）用コードブックとを入力端子123から得られるV/U判別データに応じて切換選択するような構造を有している。逆ベクトル量子化部124からの出力データは、データ数逆変換部125に送られて逆変換され、得られた振幅データが有声音合成部126及び無声音合成部127に送られる。入力端子122からの符号化ピッチデータは、ピッチ復号化部128で復号化され、データ数逆変換部125、有声音合成部126及び無声音合成部127に送られる。また入力端子123からのV/U判別データは有声音合成部126及び無声音合成部127に送られる。

＊

$$V_m(n) = A_m(n) \cos(\theta_m(n)) \quad 0 \leq n < L \quad \dots (12)$$

と表すことができる。全バンドの内のV（有声音）と判別された全てのバンドの有声音を加算（ $\sum V_m(n)$ ）して最終的な有声音V(n)を合成する。

【0068】この(12)式中の $A_m(n)$ は、上記合成フレームの先端から終端までの間で補間された第m高調波の振幅である。最も簡単には、フレーム単位で更新され

$$A_m(n) = (l-n)A_{m,0}/L + nA_{m,L}/L \quad \dots (13)$$

の式により $A_m(n)$ を計算すればよい。

★ ★【0069】次に、上記(12)式中の位相 $\theta_m(n)$ は、

$$\theta_m(n) = m\omega_{0,m}n + n^2 m(\omega_{1,m} - \omega_{0,m})/2L + \phi_{0,m} + \Delta\omega n \quad \dots (14)$$

により求めることができる。この(14)式中で、 $\phi_{0,m}$ は上記合成フレームの先端（ $n=0$ ）での第m高調波の位相（フレーム初期位相）を示し、 $\omega_{0,m}$ は合成フレーム先端（ $n=0$ ）での基本角周波数、 $\omega_{1,m}$ は該合成フレームの終端（ $n=L$ ：次の合成フレーム先端）での基本角周波数をそれぞれ示している。上記(11)式中の $\Delta\omega$ は、 $n=L$ における位相 $\phi_{1,m}$ が $\theta_m(L)$ に等しくなるような最小の $\Delta\omega$ を設定する。

【0070】以下、任意の第mバンドにおいて、それぞれ $n=0$ 、 $n=L$ のときのV/U判別結果に応じた上記振幅 $A_m(n)$ 、位相 $\theta_m(n)$ の求め方を説明する。第mバンドが、 $n=0$ 、 $n=L$ のいずれもV（有声音）と

＊【0066】有声音合成部126では例えば余弦(cosine)波合成により時間軸上の有声音波形を合成し、無声音合成部127では例えばホワイトノイズをバンドパスフィルタでフィルタリングして時間軸上の無声音波形を合成し、これらの各有声音合成波形と無声音合成波形とを加算部129で加算合成して、出力端子130より取り出すようにしている。この場合、上記振幅データ、ピッチデータ及びV/U判別データは、上記分析時の1フレーム（Lサンプル、例えば160サンプル）毎に更新されて与えられるが、フレーム間の連続性を高める（円滑化する）ために、上記振幅データやピッチデータの各値を1フレーム中の例えば中心位置における各データ値とし、次のフレームの中心位置までの間（合成時の1フレーム）の各データ値を補間により求める。すなわち、合成時の1フレーム（例えば上記分析フレームの中心から次の分析フレームの中心まで）において、先端サンプル点での各データ値と終端（次の合成フレームの先端）サンプル点での各データ値とが与えられ、これらのサンプル点間の各データ値を補間により求めるようにしている。

【0067】以下、有声音合成部126における合成処理を詳細に説明する。上記V（有声音）と判別された第mバンド（第m高調波の帯域）における時間軸上の上記1合成フレーム（Lサンプル、例えば160サンプル）分の有声音を $V_m(n)$ とすると、この合成フレーム内の時間インデックス（サンプル番号）nを用いて、

※る振幅データの第m高調波の値を直線補間すればよい。すなわち、上記合成フレームの先端（ $n=0$ ）での第m高調波の振幅値を $A_{m,0}$ 、該合成フレームの終端（ $n=L$ ：次の合成フレームの先端）での第m高調波の振幅値を $A_{m,L}$ とすると、

される場合に、振幅 $A_m(n)$ は、上述した(13)式により、伝送された振幅値 $A_{m,0}$ 、 $A_{m,L}$ を直線補間して振幅 $A_m(n)$ を算出すればよい。位相 $\theta_m(n)$ は、 $n=0$ で $\theta_m(0) = \phi_{0,m}$ から $n=L$ で $\theta_m(L)$ が $\phi_{1,m}$ となるように $\Delta\omega$ を設定する。

【0071】次に、 $n=0$ のときV（有声音）で、 $n=L$ のときUV（無声音）とされる場合に、振幅 $A_m(n)$ は、 $A_m(0)$ の伝送振幅値 $A_{m,0}$ から $A_m(L)$ で0となるように直線補間する。 $n=L$ での伝送振幅値 $A_{m,L}$ は無声音の振幅値であり、後述する無声音合成の際に用いられる。位相 $\theta_m(n)$ は、 $\theta_m(0) = \phi_{0,m}$ とし、かつ $\Delta\omega = 0$ とする。

【0072】さらに、 $n=0$ のときUV（無声音）で、 $n=L$ のときV（有声音）とされる場合には、振幅 $A_{(n)}$ は、 $n=0$ での振幅 $A_{(0)}$ を0とし、 $n=L$ で伝*

*送された振幅値 $A_{(L)}$ となるように直線補間する。位相 $\theta_{(n)}$ については、 $n=0$ での位相 $\theta_{(0)}$ として、フレーム終端での位相値 $\phi_{(L)}$ を用いて、

$$\theta_{(0)} = \phi_{(L)} - m(\omega_{01} + \omega_{L1})L/2 \quad \dots (15)$$

とし、かつ $\Delta\omega=0$ とする。

※を設定する手法について説明する。上記(14)式で、 $n=L$ と置くことにより、

【0073】上記 $n=0$ 、 $n=L$ のいずれもV（有声音）とされる場合に、 $\theta_{(L)}$ が $\phi_{(L)}$ となるように $\Delta\omega$ ※

$$\begin{aligned} \theta_{(L)} &= m\omega_{01}L + L^2 m(\omega_{L1} - \omega_{01})/2L + \phi_{0L} + \Delta\omega L \\ &= m(\omega_{01} + \omega_{L1})L/2 + \phi_{0L} + \Delta\omega L \\ &= \phi_{(L)} \end{aligned}$$

となり、これを整理すると、 $\Delta\omega$ は、

$$\Delta\omega = (\text{mod}2\pi((\phi_{(L)} - \phi_{0L}) - mL(\omega_{01} + \omega_{L1})/2)/L) \quad \dots (16)$$

となる。この(16)式で $\text{mod}2\pi(x)$ とは、 x の主値を $-\pi \sim +\pi$ の間の値で返す関数である。例えば、 $x=1.3\pi$ のとき $\text{mod}2\pi(x) = -0.7\pi$ 、 $x=2.3\pi$ のとき $\text{mod}2\pi(x) = 0.3\pi$ 、 $x=-1.3\pi$ のとき $\text{mod}2\pi(x) = 0.7\pi$ 、等である。

【0074】ここで、図16のAは、音声信号のスペクトルの一例を示しており、バンド番号（ハーモニクスナンバ） m が8、9、10の各バンドがUV（無声音）とされ、他のバンドはV（有声音）とされている。このV（有声音）のバンドの時間軸信号が上記有声音合成部126により合成され、UV（無声音）のバンドの時間軸信号が無声音合成部127で合成されるわけである。

【0075】以下、無声音合成部127における無声音合成処理を説明する。ホワイトノイズ発生部131からの時間軸上のホワイトノイズ信号波形を、所定の長さ（例えば256サンプル）で適当な窓関数（例えばハミング窓）により窓かけをし、STFT処理部132によりSTFT（ショートタイムフーリエ変換）処理を施すことにより、図16のBに示すようなホワイトノイズの周波数軸上のパワースペクトルを得る。このSTFT処理部132からのパワースペクトルをバンド振幅処理部133に送り、図16のCに示すように、上記UV（無声音）とされたバンド（例えば $m=8, 9, 10$ ）について上記振幅 $|A_{(m)}|_{uv}$ を乗算し、他のV（有声音）とされたバンドの振幅を0にする。このバンド振幅処理部133には上記振幅データ、ピッチデータ、V/UV判別データが供給されている。バンド振幅処理部133からの出力は、ISTFT処理部134に送られ、位相は元のホワイトノイズの位相を用いて逆STFT処理を施すことにより時間軸上の信号に変換する。ISTFT処理部134からの出力は、オーバーラップ加算部135に送られ、時間軸上で適当な（元の連続的なノイズ波形を復元できるように）重み付けをしながらオーバーラップ及び加算を繰り返す、連続的な時間軸波形を合成する。オーバーラップ加算部135からの出力信号が上記加算部129に送られる。

【0076】このように、各合成部126、127にお

いて合成されて時間軸上に戻された有声音部及び無声音部の各信号は、加算部129により適当な固定の混合比で加算して、出力端子130より再生された音声信号を取り出す。

【0077】なお、上記図10の音声分析側（エンコード側）の構成や図15の音声合成側（デコード側）の構成については、各部をハードウェア的に記載しているが、いわゆるDSP（デジタル信号プロセッサ）等を用いてソフトウェアプログラムにより実現することも可能である。

【0078】なお、本発明は上記実施例のみに限定されるものではなく、例えば、音声信号のみならず、音響信号を入力信号として用いることもできる。また、入力オーディオ信号（音声信号や音響信号）の特徴を表すパラメータとしては、上記V（有声音）/UV（無声音）の判別情報に限定されず、ピッチの値、ピッチ成分の強弱、信号スペクトルの傾きやレベル等を使用することができる。さらに、このような特徴パラメータは、符号化方式に応じて本来伝送するパラメータ情報の一部を代用しても、別途に伝送してもよく、他の伝送パラメータで代用させる場合には適応的コードブックと見なせ、別途伝送する場合には構造化コードブックと見なせる。

【0079】

【発明の効果】以上の説明から明らかなように、本発明に係る高能率符号化方法によれば、ベクトル量子化器に入力されたオーディオ信号（音声信号や音響信号）の状態に応じた複数種類のコードブックを設け、該入力オーディオ信号のブロック毎（フレーム毎）の特徴を表すパラメータに応じて、上記複数種類のコードブックを切り換えてベクトル量子化を施しているため、効率的なベクトル量子化が実現でき、ベクトル量子化の性能向上に役立つ。

【図面の簡単な説明】

【図1】本発明に係る高能率符号化方法が適用される符号化装置（エンコーダ）側の概略構成を示すブロック図である。

【図2】コードブックの形成（トレーニング）方法を説

明するための図である。

【図3】本発明の他の実施例の方法を説明するための符号化装置（エンコーダ）の要部の概略構成を示すブロック図である。

【図4】階層構造を有するベクトル量子化の動作を説明するための図である。

【図5】階層構造を有するベクトル量子化の動作を説明するための図である。

【図6】階層構造を有するベクトル量子化の動作を説明するための図である。

【図7】階層構造を有するベクトル量子化の動作を説明するための図である。

【図8】階層構造を有するベクトル量子化の動作を説明するための図である。

【図9】階層構造を有するベクトル量子化の動作を説明するための図である。

【図10】本発明に係る高能率符号化方法が適用される装置の具体例としての音声信号の合成分析符号化装置の分析側（エンコード側）の概略構成を示す機能ブロック図である。

【図11】窓かけ処理を説明するための図である。

【図12】窓かけ処理と窓関数との関係を説明するための図である。

【図13】直交変換（FFT）処理対象としての時間軸データを示す図である。

【図14】周波数軸上のスペクトルデータ、スペクトル包絡線（エンベロープ）及び励起信号のパワースペクトルを示す図である。

【図15】本発明に係る高能率符号化方法が適用される装置の具体例としての音声信号の合成分析符号化装置の

合成側（デコード側）の概略構成を示す機能ブロック図である。

【図16】音声信号を合成する際の無声音合成を説明するための図である。

【符号の説明】

- 12 周波数軸変換処理部
- 13 非線形圧縮部
- 14 ブロック間（フレーム間）差分処理部
- 15 ベクトル量子化部
- 15_a、25_{1a}～25_{5a} ベクトル量子化器
- 15_v、25_{1v}～25_{5v} V（有声音）用コードブック
- 15_u、25_{1u}～25_{5u} UV（無声音）用コードブック
- 15_s、25_{1s}～25_{5s} 切換スイッチ
- 21 次元低減部
- 22_a S次元ベクトル量子化器
- 22_c S次元コードブック
- 23 次元拡張部
- 103 ピッチ抽出部
- 104 窓かけ処理部
- 105 直交変換（FFT）部
- 106 高精度（ファイン）ピッチサーチ部
- 107 有声音／無声音（V／UV）判別部
- 108 振幅再評価部
- 109 データ数変換（データレートコンバート）部
- 110 ベクトル量子化部
- 126 有声音合成部
- 127 無声音合成部